UNIVERSITÉ PARIS-EST

THÈSE

Pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITÉ PARIS-EST

ÉCOLE DOCTORALE MATHEMATIQUES ET SCIENCES ET TECHNOLOGIE DE L'INFORMATION ET DE LA COMMUNICATION

DISCIPLINE : MATHÉMATIQUES

Présentée par

Pierre Youssef

Invertibilité restreinte, distance au cube et covariance de matrices aléatoires

---

Directeur de thèse : **Olivier Guédon**

Soutenue le 21 mai 2013
Devant le jury composé de

M. Frédéric Bayart	Univ. Blaise Pascal Clermont-Ferrand	Rapporteur
M. Dario Cordero-Erausquin	Univ. Pierre et Marie Curie	Examinateur
M. Matthieu Fradelizi	Univ. Paris-Est Marne-La-Vallée	Examinateur
M. Apostolos Giannopoulos	University of Athens	Rapporteur
M. Gilles Godefroy	Univ. Pierre et Marie Curie	Examinateur
M. Olivier Guédon	Univ. Paris-Est Marne-la-Vallée	Directeur de thèse
M. Philippe Jaming	Université de Bordeaux 1	Examinateur
M. Alain Pajor	Univ. Paris-Est Marne-La-Vallée	Examinateur

Thèse préparée au
Laboratoire LAMA CNRS UMR 8050
Université de Paris-Est Marne-la-Vallée
5, boulevard Descartes, Champs-sur-Marne
77454 Marne-la-Vallée cedex 2, France

Remerciement

La tradition veut que je commence par remercier mon directeur de thèse Olivier Guédon. Même en mettant de côté cette coutume, je ne vois pas comment je pourrai ne pas remercier Olivier en premier lieu. Après presque trois ans sous sa direction, je mesure la chance que j'ai de l'avoir comme directeur ; il m'a très bien orienté tout en me laissant une liberté qui, au fur et à mesure, m'a appris à me débrouiller seul. Dès le début, il a su me mettre à l'aise et m'encourager dans mes nombreuses périodes de doute. Tout au long de ma thèse, j'ai pu bénéficier de son savoir mathématique, sa disponibilité et surtout sa patience. Olivier a dû faire face à mes nombreuses demandes administratives et a dû écrire des lettres de toutes sortes pour mes multiples démarches. Je suis convaincu que si je suis en mesure de terminer ma thèse et d'avoir déjà un postdoc pour l'année prochaine, c'est en grande partie grâce à lui. Merci beaucoup Olivier, je serai toujours reconnaissant envers toi pour tout ce que tu as fait pour moi.

Frédéric Bayart a très gentiment accepté de rapporter ma thèse. J'étais agréablement surpris par sa lecture très détaillée de mon manuscrit et ses remarques qui m'ont permis d'améliorer la rédaction. Je le remercie sincèrement.

Grâce à l'école doctorale Paris-Est, j'ai pu bénéficier d'une bourse de mobilité internationale qui m'a permis d'effectuer un séjour d'un mois et demi à Athènes. C'est ainsi que j'ai rencontré Apostolos Giannopoulos. Je pense que ce voyage était un grand déclic dans ma thèse et Apostolos n'y est pas pour rien. Son hospitalité et sa générosité sans limites m'ont énormément touché. Il a accepté de rapporter ma thèse ce qui me fait énormément plaisir, je lui dit tout simplement : ευχαριστώ πάρα πολύ.

C'était ma première année en France, et j'ignorais complètement ce qu'est la recherche en Mathématiques. Par un pur hasard, je suis rentré dans le bureau de Gilles Godefroy pour lui demander de faire un travail d'étude et de recherche sous sa direction. C'est après que j'ai réalisé avoir sonné la porte d'un grand mathématicien, qui m'a beaucoup aidé et surtout su m'orienter vers un domaine qui me plaît. Il me fait l'honneur de participer à mon jury et je le remercie énormément.

Durant ma deuxième année de master, j'ai suivi deux cours de Dario Cordero-Erausquin et il faut dire que grâce à son enseignement de grande qualité, j'ai pu découvrir une branche des Mathématiques qui me plaît beaucoup. Dario a très gentiment accepté d'encadrer mon mémoire de M2 et m'a présenté à Olivier qui a également participé à l'encadrement de mon mémoire. Je suis très content qu'il ait également accepté de participer à mon jury et je le remercie du fond du coeur.

Alain Pajor a toujours été une référence par ses connaissances mathématiques et son expérience dont j'ai pu bénéficier. Matthieu Fradelizi avec sa bonne humeur était quelqu'un qui motive vraiment à travailler. Tous les deux ont accepté de faire partie de mon jury et je les remercie pour l'honneur qu'ils me font. Philippe Jaming me fait également l'honneur de participer à mon jury et je le remercie beaucoup.

Le hasard a joué un grand rôle dans tout mon parcours. Avant de quitter le Liban, je recherchais une bourse pour pouvoir poursuivre mes études en France et c'est par hasard que j'ai entendu un certain Paul (que je ne connais pas) parler de la bourse Master Ile-De-France. Il a eu la gentillesse de m'envoyer les dossiers de candidature pour cette bourse que j'ai finalement obtenue et pas lui. Même si je ne l'ai vu qu'une seule fois, c'est un peu grâce à lui que j'ai pu venir en France ; rien que pour ça, je le remercie.

Je remercie le conseil régional d'Ile-De-France avec le programme de bourse de Master qui m'a donné la possibilité de poursuivre mes études à l'université Pierre et Marie Curie. Je remercie également la France, ce pays magnifique qui m'a très bien accueilli et m'a ouvert de nouveaux horizons.

Les trois années de thèse passées au sein de l'université Paris-Est Marne-La-Vallée étaient vraiment très agréables. L'ambiance au sein du labo était très bonne, les gens sont super sympa et j'ai eu l'occasion de discuter avec plusieurs d'entre eux. J'aimerais particulièrement remercier Audrey et Christianne qui se sont occupées de plusieurs de mes démarches. Je remercie également Sylvie Cach qui a dû supporter mes demandes interminables et m'a toujours répondu avec un grand sourire. J'ai eu l'occasion durant ma thèse de faire plusieurs séjours à l'étranger, je remercie l'école doctorale Paris-Est qui a financé mon voyage en Grèce et Djalil Chafai qui avait pris en charge mon voyage aux Etats Unis. Une pensée particulière à Beatrice, Lefteris, Pantelis, Petros et Sylouanos qui ont rendu mon séjour à Athènes très agréable. Je remercie Artem Zvavitch qui m'a très bien accueilli à Kent et Mark Rudelson qui en a fait de même à Ann Arbor ; c'était un vrai honneur pour moi de les rencontrer. Je remercie également Jaegil Kim avec qui j'ai pu discuter à Kent et qui a eu la gentillesse de m'emmener à Cleveland.

J'ai régulièrement assisté au groupe de travail ASPRO de l'université Paris-Est Marne-La-Vallée ainsi qu'au séminaire d'analyse fonctionnelle et au groupe de travail Convexité et Probabilités de l'université Pierre et Marie Curie. J'ai beaucoup appris durant ces exposés et je remercie les organisateurs pour cette expérience bien réussie. J'ai également eu l'occasion de participer à plusieurs conférences pour la plupart organisées par Dario Cordero-Erausquin, Matthieu Fradelizi et Olivier Guédon ; Je retiens notamment celle à Roscoff en Juin 2012 qui m'a permis de rentrer en contact avec beaucoup de mathématiciens et de me sentir très à l'aise dans ce milieu qui regroupe des gens très sympa.

L'ambiance qui règne entre les doctorants à Marne est tout simplement magnifique. Je remercie Ali, Arnaud, Harry (notre Psy), Marie-Noemie, Marwa (notre chouchou), Omar, Paolo, Pierre (le faux), Rania, Sebastien, Victor, Xavier (le bucheron) pour tous les beaux moments et surtout les rigolades qu'il y avait au bureau. J'aimerais remercier particulièrement Bertrand ; on a commencé la thèse au même moment et on finit également en même temps. Je dois dire que j'ai vraiment eu de la chance parce qu'il m'a beaucoup aidé, tant en Latex qu'en rédaction et dans mes nombreuses questions ; il faut dire que Bertrand connaît tout sur tout. Je remercie également tous les gens que j'ai rencontrés à Paris, tous les amis que j'ai en France et que je ne vais pas énumérer par peur d'oublier quelqu'un. Une pensée particulière à Ghida qui m'a aidé dans mes nombreuses périodes de doute.

J'aimerais également remercier mon pays le Liban qui m'a offert de très beaux souvenirs que je garderai toujours en moi. Mes professeurs au collège des soeurs des saints coeurs ainsi que ceux à l'université Libanaise-

Faculté des sciences II m'ont fourni un enseignement de très bonne qualité qui m'a permis d'avoir de bonnes bases en Mathématiques. Mes amis au Liban et ceux à l'université Libanaise m'ont toujours soutenu et ont eu les mots motivants qui m'ont donné confiance en moi. Une pensée à mes tantes et mes cousins au Liban qui auraient sûrement aimé être présents pour ma soutenance.

Je garderai toujours en tête que mon cousin Sami fût la première personne que j'ai vue en arrivant à Paris. J'espère pouvoir lui faire comprendre pourquoi les ronds et les carrés sont nos amis. Ma cousine Youmna m'a toujours transmis son amour même si on ne se voyait pas souvent. Je remercie également Sonia pour les délicieux repas de Noêl et la belle ambiance qui y régnait. Un grand merci également à Viorica et sa cuisine magnifique ; j'ai toujours été très bien accueilli chez elle et Michel.

J'ai eu de la chance d'avoir deux cousins très serviables qui étaient présents quand j'avais besoin. Joseph m'a aidé pour mes nombreux déménagements et Michel a toujours été la référence pour mes nombreuses questions portant sur tout. Je n'oublierai jamais qu'ils ne m'ont jamais refusé une demande ; je les remercie du fond du coeur.

Ma tante Samia était toujours présente durant mes 5 ans passés en France. Je me souviendrai toujours comment elle s'est occupée de moi surtout durant ma première année où je venais d'arriver et je ne connaissais presque rien. Je ne saurai la remercier. Mon oncle Joseph que j'admire énormément a toujours été "le soldat de l'ombre" qui m'a aidé sans que je le sache et sans que personne le sache. J'en suis énormément reconnaissant même si je ne l'exprime pas souvent.

Je n'ai pas eu de frères par naissance mais j'ai eu la chance d'en rencontrer deux. Je remercie Karam et Karim avec qui j'ai partagé des moments inoubliables de ma vie et qui ont rendu mon séjour en France beaucoup plus agréable.

Ma plus belle découverte fut toi Kiki. Tu étais, tu es et tu resteras une grande source d'énergie positive qui me transmet l'amour de vie. Je te dis tout simplement : Merci pour ce que tu es.

Je pense que je suis béni d'avoir une famille formidable qui m'a offert toute l'affection nécessaire pour que je puisse me concentrer sur ma vie. Mes soeurs Nancy et Mimo que j'adore et avec qui j'ai une infinité de souvenirs m'ont toujours soutenu et transmis tout leur amour. Je suis sûr que sans leur présence à mes côtés, cette thèse et plein d'autres choses de ma vie ne seraient pas accomplies. Rien ne suffirait pour exprimer ma gratitude.

Ma plus grande fierté est d'avoir rendu mes parents fiers de moi. Mon père a tout fait pour m'offrir toutes les chances que lui même n'a pas eues. Je suis convaincu que s'il a eu un dixième des chances qu'il m'a offertes il aurait accomplit dix fois plus de choses que moi. Ma mère m'a démontré la puissance de l'amour ; en peu de temps, elle a appris à se servir d'un ordinateur (qu'elle maitrise maintenant) juste pour me parler sur skype. Dès que je voyage à un endroit, elle fait toutes ses recherches pour me fournir les informations du pays, où se loger, quoi visiter. Je ne pourrai exprimer toute mon admiration pour elle. Maman et Papa (boudi w badou), cette thèse vous est dédiée.

Résumé

Dans cette thèse, on aborde trois thèmes : problème de sélection de colonnes dans une matrice, distance de Banach-Mazur au cube et estimation de la covariance de matrices aléatoires. Bien que les trois thèmes paraissent éloignés, les techniques utilisées se ressemblent tout au long de la thèse.

Dans un premier lieu, nous généralisons le principe d'invertibilité restreinte de Bourgain-Tzafriri. Ce résultat permet d'extraire un "grand" bloc de colonnes linéairement indépendantes dans une matrice et d'estimer la plus petite valeur singulière de la matrice extraite. Nous proposons ensuite un algorithme déterministe pour extraire d'une matrice un bloc presque isométrique c.à.d une sous-matrice dont les valeurs singulières sont proches de 1. Ce résultat nous permet de retrouver le meilleur résultat connu sur la célèbre conjecture de Kadison-Singer. Des applications à la théorie locale des espaces de Banach ainsi qu'à l'analyse harmonique sont déduites.

Nous donnons une estimation de la distance de Banach-Mazur d'un corps convexe symétrique de $\mathbb{R}^n$ au cube de dimension n. Nous proposons une démarche plus élémentaire, basée sur le principe d'invertibilité restreinte, pour améliorer et simplifier les résultats précédents concernant ce problème.

Plusieurs travaux ont été consacrés pour approcher la matrice de covariance d'un vecteur aléatoire par la matrice de covariance empirique. Nous étendons ce problème à un cadre matriciel et nous répondons à la question. Notre résultat peut être interprété comme une quantification de la loi des grands nombres pour des matrices aléatoires symétriques semi-définies positives. L'estimation obtenue s'applique à une large classe de matrices aléatoires.

Mots clés : invertibilité restreinte, valeurs singulières, Kadison-Singer, Banach-Mazur, matrice de covariance, log-concave.

Abstract

In this thesis, we address three themes : columns subset selection inside a matrix, the Banach-Mazur distance to the cube and the estimation of the covariance of random matrices. Although the three themes seem distant, the techniques used are similar throughout the thesis.

In the first place, we generalize the restricted invertibility principle of Bougain-Tzafriri. This result allows us to extract a "large" block of linearly independent columns inside a matrix and estimate the smallest singular value of the restricted matrix. We also propose a deterministic algorithm in order to extract an almost isometric block inside a matrix i.e a submatrix whose singular values are close to 1. This result allows us to recover the best known result on the Kadison-Singer conjecture. Applications to the local theory of Banach spaces as well as to harmonic analysis are deduced.

We give an estimate of the Banach-Mazur distance between a symmetric convex body in $\mathbb{R}^n$ and the cube of dimension n. We propose an elementary approach, based on the restricted invertibility principle, in order to improve and simplify the previous results dealing with this problem.

Several studies have been devoted to approximate the covariance matrix of a random vector by its sample covariance matrix. We extend this problem to a matrix setting and we answer the question. Our result can be interpreted as a quantified law of large numbers for positive semidefinite random matrices. The estimate we obtain, applies to a large class of random matrices.

Keywords : restricted invertibility, singular values, Kadison-Singer, Banach-Mazur, covariance matrix, log-concave.

Table des matières

Introduction

Cette thèse s'inscrit dans le domaine de l'analyse fonctionnelle. Plus précisément, nous sommes intéressés par la théorie locale des espaces de Banach c.à.d leurs structures finidimensionnelles. Un espace de Banach de dimension n est comme $\mathbb{R}^n$ muni d'une norme. Comme dans chaque catégorie d'objets, on étudie les outils qui vont avec (applications, distance,..etc). Dans notre cadre, les applications sont des opérateurs linéaires donc tout simplement des matrices. Voilà pourquoi nous ajoutons un autre centre d'intérêt qui est la théorie des matrices. Les méthodes probabilistes et les phénomènes de concentration de la mesure constituent un outil puissant pour résoudre des problèmes apparaissant en théorie des espaces de Banach. Ce n'est donc pas par hasard que nous nous sommes intéressés aux matrices aléatoires. Nos résultats se situent principalement dans trois thèmes : sélection de colonnes dans une matrice, distance de Banach-Mazur au cube et estimation de la covariance de matrices aléatoires. Le fil conducteur de ces trois sujets est une méthode d'approximation de l'identité inventée par Batson-Spielman-Srivastava [13].

Théorème 0.1 (Batson-Spielman-Srivastava)**.** *Soit* $\varepsilon \in (0,1)$ *et* $m, n \in \mathbb{N}$. *Pour tout* $x_1, \dots, x_m \in \mathbb{R}^n$ *il existe* $s_1, \dots, s_m \in [0, \infty)$ *tels que*

$$\Big|\Big\{ i \in \{1, \dots, m\} : s_i \neq 0 \Big\}\Big| \leqslant \left\lceil \frac{n}{\varepsilon^2} \right\rceil, \tag{1}$$

et pour tout $y \in \mathbb{R}^n$ *on a*

$$(1-\varepsilon)^2 \sum_{i=1}^{m} \langle x_i, y\rangle^2 \leqslant \sum_{i=1}^{m} s_i \langle x_i, y\rangle^2 \leqslant (1+\varepsilon)^2 \sum_{i=1}^{m} \langle x_i, y\rangle^2. \tag{2}$$

Ce théorème nous dit que pour approcher une matrice de la forme UU^t, avec U de taille $n \times m$, le nombre de colonnes nécessaires est de l'ordre de n. Néanmoins, les colonnes choisies doivent être pondérées par des poids. Ce résultat est à la base destiné au domaine de l'informatique où la réduction du nombre de données s'avère très importante. Il vient compléter de nombreux travaux spécialement dus à Spielman et Teng ([75],[76]) qui avaient déjà démontré des résultats similaires avec des termes log parasites. Le point essentiel dans ce résultat est certainement la méthode inventée qui produit un algorithme déterministe pour effectuer l'extraction des colonnes. Elle repose sur l'étude de l'évolution des valeurs propres d'une matrice positive quand elle est perturbée par une matrice de rang 1. C'est plutôt à la méthode plus qu'au résultat même de Batson-Spielman-Srivastava [13] que l'on s'intéressera par la suite.

Notons d'abord que la première ouverture vers le domaine de la théorie locale des espaces de Banach fut l'application, due à Srivastava [77], du Théorème 0.1 au problème d'approximation d'un corps convexe par un corps convexe voisin ayant moins de points de contacts avec l'ellipsoïde de volume maximal qu'il contient. Rappelons qu'un corps convexe K est un ensemble convexe compact d'intérieur non vide ; il est symétrique si $K = -K$. Il existe une correspondance entre l'ensemble des corps convexes symétriques de $\mathbb{R}^n$ et les espaces vectoriels normés de dimension n. En effet, il suffit d'associer à un corps convexe symétrique K sa jauge $p_k(x) = \inf\{\lambda > 0/\ x \in \lambda K\}$ qui constitue une norme sur $\mathbb{R}^n$; réciproquement, étant donné une norme $\|\cdot\|$ sur $\mathbb{R}^n$, l'ensemble $\{x \in \mathbb{R}^n/\ \|x\| \leqslant 1\}$ est un corps convexe symétrique. Un résultat fondamental est celui de John [40] :

Théorème 0.2 (John). *Soit K un corps convexe de $\mathbb{R}^n$. Alors B_2^n est l'ellipsoïde de volume maximal contenu dans K si et seulement si $B_2^n \subset K$ et il existe $x_1, ..., x_m$ des points de contacts de K avec B_2^n et des scalaires $c_1, ..., c_m > 0$ tels que :*

$$Id = \sum_{j \leqslant m} c_j x_j x_j^t \quad \textit{et} \quad \sum_{j \leqslant m} c_j x_j = 0$$

Lorsque B_2^n est l'ellipsoïde de volume maximal contenu dans K, on dira que K est en position de John. Notons que le résultat précédent est encore valable lorsque B_2^n est l'ellipsoïde de volume minimal contenant K. Par la suite, plusieurs généralisations de ce résultat ont été données ([12],[30],[35],[49]). Les points de contacts jouent un rôle très important puisqu'ils caractérisent, en quelque sorte, le corps convexe associé. Le nombre de points de contacts intervenant dans la décomposition de John est inférieur à $\frac{n(n+1)}{2}$, mais rien de mieux ne pourrait être dit en toute généralité. Un problème intéressant est de réduire le nombre de points de contacts ; trouver un corps convexe L qui est proche de K c.à.d $L \subset K \subset \alpha L$, où α est une constante, tel que L a moins de points de contacts. En approchant la décomposition de l'identité associée à un corps convexe et en suivant les démarches utilisées par Rudelson [65], Srivastava a donné une première application du Théorème 0.1 au problème de réduction des points de contacts, améliorant le résultat obtenu par Rudelson [65]. A partir de là, la porte était grande ouverte vers d'autres applications de ce résultat ([33],[72],[71]). Dans cette thèse, la méthode de Batson-Spielman-Srivastava sera pertinente puisqu'elle interviendra dans les trois thèmes abordés. Commençons à présent par introduire chacun des problèmes auxquels nous nous sommes intéressés :

0.1 Invertibilité restreinte et sélection de colonnes

Soit U une matrice rectangulaire de taille $n \times m$. On verra U comme un opérateur de l_2^m dans l_2^n où l_2^n désigne $\mathbb{R}^n$ muni de la norme euclidienne :

$$\text{Pour } x = (x_j)_{j \leqslant n} \in \mathbb{R}^n, \qquad \|x\|_2 = \left(\sum_{j \leqslant n} x_j^2 \right)^{\frac{1}{2}}.$$

On notera par $s_1(U) \geqslant ... \geqslant s_{n \wedge m}(U)$ les valeurs singulières de U. La norme d'opérateur de U sera notée par $\|U\|$ et est égale à $s_1(U)$. La norme de Hilbert-Schmidt de U est donnée par

$$\|U\|_{\text{HS}} = \sqrt{\text{Tr}(UU^t)}.$$

On s'intéresse à extraire des colonnes de la matrice U de telle sorte que la matrice extraite vérifie de meilleures propriétés ou que le nombre de données nécessaires à la résolution d'un problème soit réduit.

0.1.1 Invertibilité restreinte

Une propriété essentielle des matrices est évidemment son invertibilité et plus précisément l'injectivité puisque la surjectivité peut être obtenue gratuitement en se restreignant à l'image. En général, une matrice n'est pas injective et elle ne peut l'être dans le cadre rectangulaire où la dimension de l'espace de départ est supérieure à celle de l'arrivée. Dans cette première partie, nous nous intéressons au problème suivant :

Etant donnée U une matrice de taille $n \times m$, extraire un "grand" nombre de colonnes linéairement indépendantes et estimer la plus petite valeur singulière de la matrice extraite.

Par l'algèbre linéaire élémentaire, on sait que le nombre de colonnes linéairement indépendantes que l'on peut extraire est égal au rang de la matrice. Cependant cette notion n'est pas convenable car elle n'est pas stable par petite perturbation. Pour voir ceci, prenons D une matrice diagonale de taille $n \times n$ de la forme suivante

$$D = \begin{pmatrix} 1+\delta & 0 & \dots & 0 \\ 0 & \delta & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \delta \end{pmatrix}.$$

On voit que $D - \delta \cdot Id$ est de rang 1 alors que D est de rang n. Donc après perturbation, aussi petite qu'elle soit, le rang est passé de n à 1. La notion du rang ne tenait pas compte du fait que D repose sur un seul sous-espace propre. On va remplacer la notion du rang par ce qu'on appelle le rang stable et qui est défini par

$$\mathrm{srang}(U) = \frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2}.$$

Sur l'exemple précédent, on voit que si δ est petit alors le rang stable de D est de l'ordre de 1 avant et après perturbation. Notons que

$$\mathrm{srang}(U) = \frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2} = \frac{\sum_{j=1}^{\mathrm{rang}(U)} s_j(U)^2}{s_1(U)^2} \leqslant \mathrm{rang}(U).$$

Ainsi le "grand" nombre de colonnes de U qu'on cherchera à extraire sera donné par $\mathrm{srang}(U)$. On notera à chaque fois $\sigma \subset \{1, ..., m\}$ l'ensemble des indices des colonnes choisies et U_σ la restriction de U à ces colonnes ; en d'autres termes $U_\sigma = UP_\sigma^t$ où P_σ est la projection canonique de $\mathbb{R}^n$ dans $\mathbb{R}^\sigma$. Concernant l'estimation de la plus petite valeur singulière, la question est de donner une minoration non triviale de celle ci. Il est facile de voir que

$$s_k(U) > 0 \Leftrightarrow \mathrm{rang}(U) \geqslant k,$$

ainsi donner une minoration non triviale de la plus petite valeur singulière de U_σ signifie que les colonnes de U_σ sont linéairement indépendantes ce qui répond à la question. Notons également que donner une minoration pour la plus petite valeur singulière de U_σ est équivalent à donner une majoration de la norme de l'inverse de U_σ ; en effet,

$$\forall x \in \mathbb{R}^\sigma, \quad \|U_\sigma x\|_2^2 = \langle U_\sigma x, U_\sigma x\rangle = \left\langle U_\sigma^t U_\sigma x, x\right\rangle \geqslant s_{\min}(U_\sigma)^2 \|x\|_2^2.$$

Le problème d'invertibilité restreinte a d'abord été étudié par Bourgain-Tzafriri [19] qui ont obtenu un résultat dans le cadre d'une matrice carrée ayant des colonnes de norme 1. Bourgain-Tzafriri ont démontré :

Théorème 0.3 (Bourgain-Tzafriri). *Soit T une matrice de taille $n \times n$ telle que $\|Te_j\|_2 = 1$, où $(e_j)_{j\leqslant n}$ désigne la base canonique de $\mathbb{R}^n$. Alors il existe $\sigma \subset \{1, .., n\}$ avec*

$$|\sigma| \geqslant c\frac{n}{\|T\|^2}$$

tel que pour tout choix de scalaires $(a_j)_{j\in\sigma}$

$$\left\|\sum_{j\in\sigma} a_j T e_j\right\|_2 \geqslant c' \left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}},$$

c et c' étant des constantes universelles.

Comme $\|Te_j\|_2 = 1$ pour tout $j \leqslant n$, alors $\|T\|_{\text{HS}}^2 = n$ et la taille de l'ensemble σ extrait dans le théorème précédent est proportionnelle au rang stable de T. Notons que la conclusion du théorème signifie que $s_{\min}(U_\sigma) \geqslant c'$. La preuve donnée par Bourgain-Tzafriri utilise des sélecteurs aléatoires, le lemme de Sauer-Shelah ([70], [73]) qui est un argument combinatoire et enfin le théorème de factorisation de Grothendieck (voir [27] et [61]). La démonstration est technique et non constructive. Dans [87], Tropp a proposé un algorithme probabiliste pour réaliser la factorisation de Grothendieck, ce qui lui a permis de donner un algorithme probabiliste pour effectuer l'extraction du bloc de colonnes qui fournit le résultat de Bourgain-Tzafriri. Le Théorème 0.3 est souvent connu sous le nom du principe d'invertibilité restreinte de Bourgain-Tzafriri. On peut aussi interpréter ce résultat comme l'invertibilité de l'opérateur T sur la décomposition canonique de l'identité $Id = \sum_{j\leqslant n} e_j e_j^t$; ainsi le problème revient à chercher une grande partie de cette décomposition sur laquelle T est inversible. En ce sens, Vershynin [90] a généralisé ce résultat pour une décomposition quelconque de l'identité améliorant par la même occasion la taille de la partie extraite. Notons que Vershynin a également démontré une estimation de la norme de l'opérateur restreint (et non seulement de celle de l'inverse) ; nous préférons à présent faire appel uniquement à une partie du résultat obtenu par Vershynin, celle en relation avec le principe d'invertibilité restreinte, sachant que le résultat complet fera l'objet d'une étude plus détaillée dans la section suivante.

Théorème 0.4 (Vershynin)**.** *Soit* $Id = \sum_{j\leqslant m} x_j x_j^t$ *une décomposition de l'identité sur* $\mathbb{R}^n$ *et soit* T *un opérateur linéaire sur* l_2^n*. Pour tout* $\varepsilon \in (0,1)$ *il existe* $\sigma \subset \{1,\ldots,m\}$ *avec*

$$|\sigma| \geqslant (1-\varepsilon)\text{srang}(T) = (1-\varepsilon)\frac{\|T\|_{\text{HS}}^2}{\|T\|^2}$$

tel que pour toute famille de scalaires $(a_j)_{j\in\sigma}$

$$\left\|\sum_{j\in\sigma} a_j \frac{Tx_j}{\|Tx_j\|_2}\right\|_2 \geqslant c(\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}},$$

où $c(\varepsilon)$ *est une constante qui ne dépend que de* ε*.*

Commencons d'abord par comparer ce résultat avec le précédent. Si, à la place d'une décomposition quelconque de l'identité, on prenait la décomposition canonique de l'identité et qu'on supposait en plus que $\|Te_j\|_2 = 1$ alors le Théorème 0.4 permet de trouver σ de taille $(1-\varepsilon)\frac{n}{\|T\|^2}$ tel que

$$\left\|\sum_{j\in\sigma} a_j Te_j\right\|_2 \geqslant c(\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}},$$

pour toute suite de scalaires $(a_j)_{j\in\sigma}$. On retrouve donc le Théorème 0.3 avec l'avantage de pouvoir extraire un ensemble d'indices de taille presque égal au rang stable de T, alors qu'avant on ne pouvait pas extraire plus qu'une proportion du rang stable. Ceci s'avère très important, car pour plusieurs applications (on en verra quelques unes par la suite), on a besoin presque de la totalité et une proportion n'est pas suffisante. La constante dépendant de ε joue également un rôle crucial et trouver la bonne dépendance en ε est au coeur du problème. Notons également que dans ce résultat, Vershynin extrait des vecteurs Tx_j mais l'estimation concerne les vecteurs normalisés ; cette normalisation sera essentielle pour certaines applications surtout lorsqu'il s'agira d'estimer la distance de Banch-Mazur au cube comme on le verra dans la deuxième partie. Concernant la preuve, Vershynin démontre son résultat par une itération assez technique du Théorème 0.3 combiné à un résultat de Kashin-Tzafriri [43] dont on discutera plus tard. Ainsi, la preuve proposée par Vershynin n'est pas constructive.

Spielman-Srivastava [74] ont également généralisé le principe d'invertibilité restreinte de Bourgain-Tzafriri. En se basant sur la méthode introduite par Batson-Spielman-Srivastava [13], ils ont proposé un algorithme déterministe pour extraire le bloc réalisant l'invertibilité.

Théorème 0.5 (Spielman-Srivastava). *Soit* $Id = \sum_{i \leqslant m} x_i x_i^t$ *une décomposition de l'identité sur* $\mathbb{R}^n$ *et soit* $\varepsilon \in (0,1)$*. Pour tout opérateur linéaire* $T : \ell_2^n \to \ell_2^n$ *il existe* $\sigma \subset \{1, \ldots, m\}$ *de taille* $|\sigma| \geqslant \left\lfloor (1-\varepsilon)^2 \frac{\|T\|_{\mathrm{HS}}^2}{\|T\|_2^2} \right\rfloor$ *tels que* $\{Tx_i\}_{i\in\sigma}$ *sont linéairement indépendants et*

$$\lambda_{\min}\left(\sum_{i\in\sigma}(Tx_i)(Tx_i)^t\right) > \frac{\varepsilon^2\|T\|_{\mathrm{HS}}^2}{m},$$

où $\lambda_{\min}$ *est calculée sur* $\mathrm{vect}\{Tx_i\}_{i\in\sigma}$.

De manière équivalente, pour toute suite de scalaires $(a_j)_{j\in\sigma}$

$$\left\|\sum_{j\in\sigma} a_j T x_j\right\|_2 \geqslant \varepsilon \frac{\|T\|_{\mathrm{HS}}}{\sqrt{m}} \left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}.$$

Ce résultat peut être interprété comme l'invertibilité restreinte de matrices rectangulaires. En effet, étant donné $Id = \sum_{i\leqslant m} x_i x_i^t$ une décomposition de l'identité sur $\mathbb{R}^n$ et T un opérateur linéaire sur l_2^n, on peut leur associer une matrice U de taille $n \times m$ ayant pour colonnes les vecteurs $(Tx_j)_{j\leqslant m}$. Puisque les $(x_j)_{j\leqslant m}$ forment une décomposition de l'identité, il est assez facile de voir que $UU^t = TT^t$ et ainsi que $\|U\| = \|T\|$ et $\|U\|_{\mathrm{HS}} = \|T\|_{\mathrm{HS}}$. Le résultat précédent peut donc être exprimé en fonction de la matrice rectangulaire U.

En conclusion, on est en face de deux résultats qui généralisent le principe d'invertibilité restreinte ; d'un côté le résultat de Vershynin où les vecteurs choisis sont normalisés mais où la dépendance en ε n'est pas satisfaisante, et d'un autre côté le résultat de Spielman-Srivastava qui fournit une bonne dépendance en ε mais où les vecteurs choisis ne sont pas normalisés. Une question naturelle est donc de trouver un résultat qui englobe les deux précédents pour disposer des normalisations tout en gardant une bonne dépendance en ε.

On montre un principe d'invertibilité restreinte pour toute matrice rectangulaire et pour tout choix de normalisations des colonnes tout en gardant une bonne dépendance en ε.

Théorème 1. *Soient* U *une matrice de taille* $n \times m$ *et* D *une matrice diagonale de taille* $m \times m$ *ayant* $(\alpha_j)_{j\leqslant m}$ *sur sa diagonale. Si* $\mathrm{Ker}(D) \subset \mathrm{Ker}(U)$*, alors pour tout* $\varepsilon \in (0,1)$ *il existe* $\sigma \subset \{1, \ldots, m\}$ *avec*

$$|\sigma| \geqslant (1-\varepsilon)^2 \mathrm{srang}(U) = (1-\varepsilon)^2 \frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2}$$

tel que

$$s_{\min}\left(U_\sigma D_\sigma^{-1}\right) > \frac{\varepsilon\|U\|_{\mathrm{HS}}}{\|D\|_{\mathrm{HS}}},$$

où $s_{\min}$ *désigne la plus petite valeur singulière.*

De manière équivalente, pour toute suite de scalaires $(a_j)_{j\in\sigma}$

$$\left\|\sum_{j\in\sigma} a_j \frac{Ue_j}{\alpha_j}\right\|_2 \geqslant \varepsilon \frac{\|U\|_{\mathrm{HS}}}{\|D\|_{\mathrm{HS}}} \left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}.$$

Pour être convaincu que le Théorème 1 n'est pas qu'une simple conséquence du Théorème 0.5, il suffit de remarquer que le point important dans le résultat précédent est que la taille de l'extraction ne dépend que de la matrice U et non de la matrice de normalisation D. Le Théorème 1 fait le pont entre les résultats précédents et généralise Théorème 0.4 et Théorème 0.5. Soient $Id = \sum_{j\leqslant m} x_j x_j^t$ une décomposition de l'identité sur $\mathbb{R}^n$ et T un opérateur linéaire sur l_2^n. Définissons U la matrice de taille $n \times m$ ayant les $(Tx_j)_{j\leqslant m}$ comme colonnes. Ainsi, comme on l'a déjà fait remarquer précédemment, on a $UU^t = TT^t$, et en appliquant le Théorème 1 on trouve $\sigma \subset \{1, ..., m\}$ de taille

$$|\sigma| \geqslant (1-\varepsilon)^2 \mathrm{srang}(U) = (1-\varepsilon)^2 \mathrm{srang}(T)$$

En prenant $D = Id$, on obtient que pour toute suite de scalaires $(a_j)_{j\in\sigma}$

$$\left\|\sum_{j\in\sigma} a_j Tx_j\right\|_2 = \left\|\sum_{j\in\sigma} a_j Ue_j\right\|_2 \geqslant \varepsilon \frac{\|U\|_{\mathrm{HS}}}{\|D\|_{\mathrm{HS}}} \left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}} = \varepsilon \frac{\|T\|_{\mathrm{HS}}}{\sqrt{m}} \left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}},$$

et on retrouve le Théorème 0.5.

En prenant $D = diag(\alpha_1, ..., \alpha_m)$ avec $\alpha_j = \|Tx_j\|_2$, on obtient que pour toute suite de scalaires $(a_j)_{j\in\sigma}$

$$\left\|\sum_{j\in\sigma} a_j \frac{Tx_j}{\|Tx_j\|_2}\right\|_2 = \left\|\sum_{j\in\sigma} a_j \frac{Ue_j}{\alpha_j}\right\|_2 \geqslant \varepsilon \frac{\|U\|_{\mathrm{HS}}}{\|D\|_{\mathrm{HS}}} \left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}} = \varepsilon \left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}},$$

et on retrouve le Théorème 0.4.

0.1.2 Norme de matrice restreinte

Dans ce qui précède, on a étudié le problème d'invertibilité qui concerne uniquement la plus petite valeur singulière. Dans cette partie, on ne s'occupera que de la plus grande valeur singulière qui représente la norme de la matrice. Etant donnés U une matrice rectangulaire de taille $n \times m$ et un entier $k \leqslant m$, on s'intéresse à extraire de U une matrice à k colonnes qui minimise la norme d'opérateur. Cette question a d'abord été abordée dans [51] puis par Kashin-Tzafriri [43]. Kashin-Tzafriri ([43], voir aussi [90]) ont démontré ce qui suit :

Théorème 0.6 (Kashin-Tzafriri). *Soit U une matrice de taille $n \times m$. Si λ est tel que $1/m \leqslant \lambda \leqslant \frac{1}{4}$, alors il existe $\sigma \subset \{1, \dots, m\}$ de taille $|\sigma| \geqslant \lambda m$ tel que*

$$\|U_\sigma\| \leqslant c \left(\sqrt{\lambda} \|U\| + \frac{\|U\|_{\mathrm{HS}}}{\sqrt{m}} \right),$$

où c est une constante universelle.

La conclusion du théorème signifie que pour $\lambda \leqslant \frac{1}{4}$ fixé on a

$$\min_{\substack{\sigma \subset \{1,\dots,m\} \\ |\sigma| = \lambda m}} \|U_\sigma\| \leqslant c \left(\sqrt{\lambda} \|U\| + \frac{\|U\|_{\mathrm{HS}}}{\sqrt{m}} \right) \tag{3}$$

Ce résultat est optimal (à constante près) dans le sens où le membre de droite dans l'estimation ne pourrait pas être remplacé par une quantité plus petite. Pour en être convaincu, considérons l'exemple suivant :

Soient $m, h, k \in \mathbb{N}$ tels que $m = hk$. On divise l'ensemble $\{1, ..., m\}$ en h ensembles disjoints I_l de taille k chacun. Soit U la matrice de taille $m \times m$ définie par $Ue_j = \frac{1}{\sqrt{k}} e_l$ si $j \in I_l$ pour $l \leqslant h$, $(e_j)_{j \leqslant m}$ étant la base canonique de $\mathbb{R}^m$. On a donc

$$UU^t = \sum_{j \leqslant m} (Ue_j)(Ue_j)^t = \sum_{l \leqslant h} \sum_{j \in I_l} \frac{1}{k} e_l e_l^t = \sum_{l \leqslant h} e_l e_l^t$$

Ainsi $\|U\| = 1$ et $\|U\|_{\mathrm{HS}} = \sqrt{h}$. Soit maintenant p le nombre de colonnes qu'on souhaite extraire (p joue le rôle de λm dans le Théorème 0.6). Si $p \leqslant h$, alors la meilleure restriction serait de choisir un vecteur de chaque bloc et dans ce cas la norme de la restriction serait $\frac{1}{\sqrt{k}} = \sqrt{\frac{h}{m}}$ qui est le deuxième terme dans (3). Si $p > h$ alors la meilleure restriction serait de choisir le plus de vecteurs dans des blocs différents, ce qui revient à choisir $\frac{p}{h}$ vecteurs de chaque bloc. Dans ce cas, la norme de la restriction est $\sqrt{\frac{p}{hk}} = \sqrt{\frac{p}{m}}$ qui joue le rôle de $\sqrt{\lambda}$ figurant comme premier terme dans (3).

La preuve du Théorème 0.6 ([43], voir aussi [90]) utilise des sélecteurs aléatoires et le théorème de factorisation de Grothendieck (voir [27] et [61]). La méthode n'est donc pas constructive. Dans [87], Tropp a donné un algorithme probabiliste pour réaliser la factorisation de Grothendieck ce qui lui a permis de donner un algorithme probabiliste pour trouver l'ensemble σ garanti par le Théorème 0.6. Notre but est de donner une nouvelle preuve de ce résultat en s'inspirant de la méthode introduite par Batson-Spielman-Srivastava [13].

Nous obtenons un résultat qui améliore la taille de la matrice extraite et toutes les constantes apparaissant dans le Théorème 0.6. Un point important est que notre preuve fournit un algorithme déterministe pour réaliser l'extraction.

Théorème 2. *Soit U une matrice de taille $n \times m$. Si $0 < \lambda \leqslant \eta < 1$, alors il existe $\sigma \subset \{1, \dots, m\}$ de taille $|\sigma| \geqslant \lambda m$ tel que*

$$\|U_\sigma\| \leqslant \frac{1}{\sqrt{1-\lambda}} \left(\sqrt{\lambda + \eta} \|U\| + \sqrt{1 + \frac{\lambda}{\eta}} \frac{\|U\|_{\mathrm{HS}}}{\sqrt{m}} \right),$$

En particulier,

$$\|U_\sigma\| \leqslant \frac{\sqrt{2}}{\sqrt{1-\lambda}} \left(\sqrt{\lambda} \|U\| + \frac{\|U\|_{\mathrm{HS}}}{\sqrt{m}} \right)$$

0.1.3 Extraction d'un bloc carré : première tentative

Jusqu'à présent, on s'est contenté d'extraire un bloc de colonnes de la matrice de départ. Etant donnée T une matrice carré de taille $n \times n$ à diagonale nulle, Bourgain-Tzafriri ([19],[20]) ont cherché à extraire une "grande" sous-matrice carré dont la norme est petite. Ils ont démontré le résultat suivant :

Théorème 0.7. *Soit T une matrice de taille $n \times n$ tel que $\langle Te_i, e_i \rangle = 0$ pour tout $i \in \{1, \dots, n\}$. Pour tout $\varepsilon \in (0,1)$, il existe $\sigma \subseteq \{1, \dots, n\}$ de taille $|\sigma| \geqslant c\varepsilon^2 n$ telle que*

$$\|P_\sigma T P_\sigma^t\| \leqslant \varepsilon \|T\|,$$

P_σ étant la projection canonique de $\mathbb{R}^n$ dans $\mathbb{R}^\sigma$ et c une constante universelle.

Il faut noter que le Théorème 0.7 implique le principe d'invertibilité restreinte de Bourgain-Tzafriri (Théorème 0.3) avec cependant une mauvaise dépendance en la norme de T. Pour voir ceci, prenons T une matrice de taille $n \times n$ tels que $\|Te_j\|_2 = 1$ pour tout $j \leqslant n$. On pose $A = T^t T - Id$, alors A est à diagonale nulle. On applique le Théorème 0.7 pour trouver $\sigma \subset \{1, ..., n\}$ de taille $c\varepsilon^2 n$ tels que $\|P_\sigma A P_\sigma^t\| \leqslant \varepsilon \|A\|$. Ceci signifie que

$$-\varepsilon \|A\| \cdot Id \preceq P_\sigma A P_\sigma^t \preceq \varepsilon \|A\| \cdot Id$$

En remarquant que $\|T\| \geqslant 1$ et donc que $\|A\| \leqslant 2\|T\|^2$, on a

$$(1 - 2\varepsilon\|T\|^2) \cdot Id \preceq P_\sigma T^t T P_\sigma^t \preceq (1 + 2\varepsilon\|T\|^2) \cdot Id$$

En prenant $\varepsilon = \frac{1}{4\|T\|^2}$, on a trouvé σ de taille $c\frac{n}{\|T\|^4}$ tels que $s_{\min}(T_\sigma) \geqslant \frac{1}{2}$ et on retrouve ainsi le Théorème 0.3 avec une mauvaise dépendance en la norme de T.

Dans [14], il a été démontré que la dépendance quadratique en ε dans le Théorème 0.7 est optimale. Pour voir ceci, prenons A une matrice d'Hadamard de taille $n \times n$ c.à.d une matrice dont les colonnes sont formées par des 1 et des -1 et sont orthogonales deux à deux. On a $AA^t = nId$, et ainsi $\|A\| = \sqrt{n}$. Si P est une projection canonique de rang k c.à.d P est une matrice diagonale ayant k termes égaux à 1 sur sa diagonale, alors PAP^t est une matrice de taille $k \times k$ ayant que des 1 ou -1. Donc $\|PAP^t\| \geqslant \sqrt{k}$ et on a

$$\|P\,(A - \mathrm{diag}(A))\,P^t\| \geqslant \|PAP^t\| - \|P\mathrm{diag}(A)P^t\| \geqslant \sqrt{k} - 1 \geqslant \frac{\sqrt{k}}{2},$$

si $k \geqslant 4$ par exemple. En conclusion, si on note $B = \frac{1}{\sqrt{n}}\,(A - \mathrm{diag}(A))$ où A est une matrice d'Hadamard de taille $n \times n$ alors pour tout ensemble d'indices σ de taille $k = \varepsilon^2 n$, on vient de voir que

$$\|P_\sigma B P_\sigma^t\| \geqslant \frac{\sqrt{k}}{2\sqrt{n}} = \frac{\varepsilon}{2} \geqslant \frac{\varepsilon}{4}\|B\|,$$

car $\|B\| \leqslant 2$. Ceci montre que la dépendance quadratique en ε dans le Théorème 0.7 est nécessaire.

Dans un survey [55] sur la méthode de Batson-Spielman-Srivastava [13], Naor a posé la question de donner une nouvelle preuve du Théorème 0.7 en utilisant les méthodes introduites par Batson-Spielman-Srivastava. Nous nous sommes donc intéressés à ce problème et on présente pour commencer une première tentative qui ne donnera pas la bonne dépendance en ε.
Si T est une matrice symétrique de taille $n \times n$, on pose $U = \left[\frac{1}{\|T\|}\,(T + \|T\| \cdot Id)\right]^{\frac{1}{2}}$. Pour montrer que $\|P_\sigma T P_\sigma^t\| \leqslant \varepsilon\|T\|$, il nous suffit de montrer que

$$(1 - \varepsilon) \cdot Id \preceq P_\sigma U^t U P_\sigma^t \preceq (1 + \varepsilon) \cdot Id$$

L'idée est de lancer l'algorithme du Théorème 2 pour trouver un ensemble d'indices ν qui vérifie la majoration, puis lancer l'algorithme du Théorème 1 pour trouver l'ensemble d'indices σ à l'intérieur de ν et qui vérifie la minoration. Cette démarche nous permet de montrer ce qui suit :

Proposition 1. *Soit T une matrice symétrique de taille $n \times n$ à diagonale nulle. Pour tout $\varepsilon \in (0,1)$, il existe $\sigma \subset \{1,..,n\}$ de taille $c\varepsilon^4 n$ tels que $\|P_\sigma T P_\sigma^t\| \leqslant \varepsilon \|T\|$.*

Ainsi notre première tentative fournit un algorithme déterministe pour montrer le Théorème 0.7 avec cependant une mauvaise dépendance en ε et l'hypothèse de symétrie sur la matrice T. Le fait de lancer chaque algorithme séparément produit cette erreur dans la dépendance en ε. Nous nous intéresserons par la suite à trouver une méthode permettant d'obtenir une majoration et une minoration des valeurs singulières de la matrice extraite.

0.1.4 Sélection d'un bloc bien conditionné

Dans la première partie, nous avions évoqué une partie d'un résultat de Vershynin [90] qu'on a énoncée dans le Théorème 0.4. L'énoncé complet est le suivant :

Théorème 0.8. *Soit $Id = \sum_{j \leqslant m} x_j x_j^t$ une décomposition de l'identité sur $\mathbb{R}^n$ et soit T un opérateur linéaire sur l_2^n. Pour tout $\varepsilon \in (0,1)$ il existe $\sigma \subset \{1,\dots,m\}$ de taille*

$$|\sigma| \geqslant (1-\varepsilon)\text{srang}(T) = (1-\varepsilon)\frac{\|T\|_{\text{HS}}^2}{\|T\|^2}$$

tels que pour toute famille de scalaires $(a_j)_{j\in\sigma}$

$$c_1(\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}} \leqslant \left\|\sum_{j\in\sigma} a_j \frac{Tx_j}{\|Tx_j\|_2}\right\|_2 \leqslant c_2(\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}},$$

où $c_1(\varepsilon)$ et $c_2(\varepsilon)$ ne dépendent que de ε.

La dépendence en ε vérifie $c(\varepsilon) = \frac{c_2(\varepsilon)}{c_1(\varepsilon)} \approx \varepsilon^{c\log(\varepsilon)}$. En d'autres termes, le résultat précédent affirme qu'on peut trouver une "grande" partie (disons de taille k) de la suite $(Tx_j)_{j\leqslant m}$ qui est $c(\varepsilon)$-équivalente à une base orthogonale de l_2^k. Vershynin applique ce résultat à l'étude des points de contacts et des plongements du cube dans un espace de Banach quelconque ; en effet, son résultat combiné à des résultats de Talagrand [83] donne une idée plus claire sur la forme de ces plongements. Dans le Théorème 1, nous avons amélioré une partie du Théorème 0.8 ; précisément, la minoration qui concerne le principe d'invertibilité restreinte. Clairement, notre but est d'obtenir une amélioration du résultat complet de Vershynin. Commençons par traduire ce résultat dans un langage matriciel :

Etant donnés $Id = \sum_{j \leqslant m} x_j x_j^t$ une décomposition de l'identité sur $\mathbb{R}^n$ et T un opérateur linéaire sur l_2^n, on considère la matrice U de taille $n \times m$ ayant les vecteurs $(Tx_j)_{j \leqslant m}$ comme colonnes. On note $\tilde{U}$ la matrice U avec des colonnes normalisées. Il est clair que $UU^t = TT^t$, et ainsi le Théorème 0.8 implique l'existence d'un ensemble d'indices σ de taille $(1-\varepsilon)\text{srang}(U)$ tel que

$$c_1(\varepsilon) \leqslant s_{\min}(\tilde{U}_\sigma) \leqslant s_{\max}(\tilde{U}_\sigma) \leqslant c_2(\varepsilon).$$

Le nombre de conditionnement d'une matrice U est donné par

$$\kappa(U) = \max\left\{ \frac{\|Ux\|_2}{\|Uy\|_2};\ \|x\|_2 = \|y\|_2 = 1 \right\} = \frac{s_{\max}(U)}{s_{\min}(U)}$$

Evidemment, si la matrice n'est pas de rang complet alors son nombre de conditionnement explose. Un problème intéressant serait d'extraire un grand bloc de la matrice tels que le nombre de conditionnement de la restriction est bien majoré. Si le nombre de conditionnement est proche de 1, alors la matrice est un multiple d'une isométrie. Le Théorème 0.8 permet d'extraire de $\tilde{U}$ un nombre de colonnes égal à $(1-\varepsilon)\text{srang}(U)$ tels que $\kappa(\tilde{U}_\sigma) \leqslant \varepsilon^{c \log(\varepsilon)}$.

Pour améliorer le résultat obtenu par Vershynin, l'idée est de fusionner les algorithmes du Théorème 1 et du Théorème 2 pour obtenir à la fois un principe d'invertibilité restreinte (et donc une estimation de $s_{\min}$) et une estimation de $s_{\max}$. Nous avons réussi à démontrer ce qui suit :

Théorème 3. *Soit U une matice de taille $n \times m$. On note $\tilde{U}$ la matrice dont les colonnes sont celles de U normalisées. Pour tout $\varepsilon \in (0,1)$, il existe $\sigma \subset \{1, ..., m\}$ de taille*

$$|\sigma| \geqslant (1-\varepsilon)^2 \text{srang}(U) = (1-\varepsilon)^2 \frac{\|U\|_{\text{HS}}^2}{\|U\|^2}$$

tel que

$$\frac{\varepsilon}{2-\varepsilon} \leqslant s_{\min}\left(\tilde{U}_\sigma\right) \leqslant s_{\max}\left(\tilde{U}_\sigma\right) \leqslant \frac{2-\varepsilon}{\varepsilon}$$

De manière équivalente, pour toute suite de scalaires $(a_j)_{j\in\sigma}$

$$\frac{\varepsilon}{2-\varepsilon}\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}} \leqslant \left\| \sum_{j\in\sigma} a_j \frac{Ue_j}{\|Ue_j\|_2} \right\|_2 \leqslant \frac{2-\varepsilon}{\varepsilon}\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}.$$

Ainsi notre résultat implique l'existence d'un bloc de taille $(1-\varepsilon)^2\text{srang}(U)$ qui a un nombre de conditionnement inférieur à $\left(\frac{2-\varepsilon}{\varepsilon}\right)^2$. Dans le régime où ε est proche de 1, le Théorème 3 nous permet d'extraire un bloc presque isométrique.

Corollaire 1. *Soit U une matrice de taille $n \times m$. On note $\widetilde{U}$ la matrice dont les colonnes sont celles de U normalisées. Pour tout $\varepsilon \in (0,1)$, il existe $\sigma \subset \{1, ..., m\}$ de taille*

$$|\sigma| \geqslant \frac{\varepsilon^2}{9}\mathrm{srang}(U) = \frac{\varepsilon^2}{9}\cdot\frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2}$$

tels que

$$1-\varepsilon \leqslant s_{\min}\left(\widetilde{U}_\sigma\right) \leqslant s_{\max}\left(\widetilde{U}_\sigma\right) \leqslant 1+\varepsilon$$

De manière équivalente, pour toute suite de scalaires $(a_j)_{j\in\sigma}$

$$(1-\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}} \leqslant \left\|\sum_{j\in\sigma} a_j \frac{Ue_j}{\|Ue_j\|_2}\right\|_2 \leqslant (1+\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}$$

On verra par la suite que le Corollaire 1 implique (dans le cas d'un opérateur symétrique) le Théorème 0.7 ce qui répond à la question de Naor. Comme on a vu que la dépendance en ε dans le Théorème 0.7 est optimale, cela veut dire qu'à constante près, le Corollaire 1 est optimal.

0.1.5 Pavage de colonnes et conjecture de Kadison-Singer

Une matrice rectangulaire U de taille $n \times m$ est dite standard si toutes ses colonnes sont de norme 1. Remarquons que le principe d'invertibilité restreinte de Bourgain-Tzafriri (Théorème 0.3) concernait des matrices standards. Le problème de pavage de colonnes consiste à partitionner la matrice en blocs bien conditionnés. Précisément, on souhaite donner un algorithme qui permet de partitionner une matrice en blocs presque isométriques. Un premier résultat dans cette direction est dû à Bourgain-Tzafriri ([19],[20]) mais n'est pas constructif. Tropp a proposé un algorithme aléatoire pour effectuer le pavage en blocs presque isométriques (voir [86] pour plus d'informations sur ce sujet).

En utilisant le Corollaire 1, on donne un algorithme déterministe pour retrouver un résultat de Bourgain-Tzafriri et effectuer le pavage en améliorant toutes les constantes qui interviennent.

Proposition 2. *Soit U une matrice standard de taille $n \times m$. Pour tout $\varepsilon \in (0,1)$, il existe une partition de $\{1, ..., m\}$ en p ensembles $\sigma_1, ..., \sigma_p$ tels que*

$$p \leqslant \frac{9\|U\|^2 \log(m)}{\varepsilon^2}$$

et pour tout $i \leqslant p$,

$$1-\varepsilon \leqslant s_{\min}\left(U_{\sigma_i}\right) \leqslant s_{\max}\left(U_{\sigma_i}\right) \leqslant 1+\varepsilon$$

Revenons à présent au problème d'extraction de matrice carré. Dans notre première tentative, nous avions réussi à retrouver le Théorème 0.7 avec cependant une mauvaise dépendance en ε. Le fait d'avoir lancé les algorithmes du Théorème 2 et du Théorème 1 consécutivement était la cause de cette erreur d'estimation. Maintenant qu'on a réussi à donner un algorithme fusionnant les deux autres, on retrouve le Théorème 0.7 avec la bonne dépendance en ε mais toujours dans le cadre des matrices symétriques. En effet, si T est une matrice symétrique de taille $n \times n$, on pose $U = \left[\frac{1}{\|T\|}(T + \|T\| \cdot Id)\right]^{\frac{1}{2}}$ et ainsi pour montrer que $\|P_\sigma T P_\sigma^t\| \leqslant \varepsilon \|T\|$, il nous suffit de montrer que

$$(1-\varepsilon) \cdot Id \preceq P_\sigma U^t U P_\sigma^t \preceq (1+\varepsilon) \cdot Id,$$

ce qui a déjà été démontré dans le Corollaire 1.

En suivant la démarche décrite précédemment, nous obtenons :

Proposition 3. *Soit T une matrice symétrique de taille $n \times n$ à diagonale nulle. Pour tout $\varepsilon \in (0,1)$, il existe $\sigma \subset \{1, ..., n\}$ de taille*

$$|\sigma| \geqslant \frac{(\sqrt{2}-1)^4 \varepsilon^2 n}{2}$$

tel que

$$\|P_\sigma T P_\sigma^*\| \leqslant \varepsilon \|T\|.$$

Le résultat précédent est important car il est directement lié à la conjecture de Kadison-Singer [41]. Cette conjecture est toujours non résolue et plusieurs formulations équivalentes ont été proposées ([7], voir [22] et [23] pour plus de détails). Nous présentons ici ce qu'on appelle la "Paving conjecture" due à Anderson [7] et qui est équivalente à Kadison-Singer.

Conjecture. *Pour tout $\varepsilon > 0$, il existe $p = p(\varepsilon)$ tels que pour tout $n \in \mathbb{N}$ et toute matrice T de taille $n \times n$ à diagonale nulle, il existe une partition de $\{1, ..., n\}$ en p ensembles $\sigma_1, ..., \sigma_p$ tels que*

$$\forall i \leqslant p, \quad \|P_{\sigma_i} T P_{\sigma_i}^t\| \leqslant \varepsilon \|T\|.$$

Notons qu'il suffirait de montrer cette conjecture dans le cadre des matrices symétriques (voir [22]). Les deux meilleurs résultats connus sur cette conjecture sont dus à Bourgain-Tzafriri [20]. Le premier montre en toute généralité l'existence d'une partition de taille de l'ordre de $\log(n)$

pour laquelle la conclusion de la conjecture est vraie. Le deuxième affirme que la conjecture est vraie pour les matrices ayant des entrées uniformément bornées par la bonne quantité.

En itérant la Proposition 3, on donne un algorithme déterministe, qui améliore les constantes intervenant dans le résultat, permettant de partitionner une matrice symétrique à diagonale nulle en un nombre de blocs de l'ordre de $\log(n)$ pour lequel la conclusion de la conjecture est vraie.

Proposition 4. *Soit T une matrice symétrique de taille $n \times n$ à diagonale nulle. Pour tout $\varepsilon \in (0,1)$, il existe une partition de $\{1, ..., n\}$ en k ensembles $\sigma_1, .., \sigma_k$ tels que*

$$k \leqslant \frac{2\log(n)}{(\sqrt{2}-1)^4 \varepsilon^2}$$

et pour tout $i \leqslant k$,

$$\left\| P_{\sigma_i} T P_{\sigma_i}^t \right\| \leqslant \varepsilon \|T\|$$

0.1.6 Application à l'analyse harmonique

Notons $\mathbb{T}$ le cercle unité et ν la mesure de Lebesgue normalisée sur $\mathbb{T}$. L'ensemble des fonctions de carré intégrable sur $\mathbb{T}$ est noté par $L_2(\mathbb{T}, \nu)$. Pour tout $f \in L_2(\mathbb{T}, \nu)$, on définit

$$\|f\|_{L_2(\mathbb{T})} = \left(\int_{\mathbb{T}} |f|^2 d\nu \right)^{\frac{1}{2}}.$$

Pour tout sous-ensemble B de $\mathbb{T}$, on définit

$$\|f\|_{L_2(B)} = \left(\frac{1}{\nu(B)} \int_{\mathbb{B}} |f|^2 d\nu \right)^{\frac{1}{2}}.$$

Si Λ est un ensemble d'entiers, la densité de Λ est donnée par :

$$\text{dens}(\Lambda) = \lim_{n \to \infty} \frac{|\Lambda \cap \{-n, n\}|}{2n},$$

pourvue que cette limite existe.
On note également $L_2^\Lambda(\mathbb{T}, \nu)$ le sous-espace de $L_2(\mathbb{T}, \nu)$ engendré par $\{e^{i.kx}\}_{k \in \Lambda}$. En d'autres termes, $L_2^\Lambda(\mathbb{T}, \nu)$ constitue l'ensemble des fonctions telles que le support de leur transformée de Fourier est inclus dans Λ.

Dans [19], Bourgain-Tzafriri ont donné une application du principe d'invertibilité restreinte au domaine de l'analyse harmonique en montrant :

Théorème 0.9 (Bourgain-Tzafriri). *Pour tout $B \subset \mathbb{T}$, il existe Λ un ensemble d'entiers avec* $\mathrm{dens}(\Lambda) \geqslant c\nu(B)$, *tel que pour tout $f \in L_2^\Lambda(\mathbb{T}, \nu)$, on a*

$$\|f\|_{L_2(B)} \geqslant c' \cdot \|f\|_{L_2(\mathbb{T})}, \tag{4}$$

où c, c' sont des constantes universelles.

Dans [89], Vershynin utilise le Théorème 0.8 pour donner également une majoration dans (4). En utilisant le Théorème 3, nous améliorons les résultats obtenus par Bourgain-Tzafriri [19] et Vershynin [89].

Théorème 4. *Soit $B \subset \mathbb{T}$ tels que $\nu(B) > 0$. Pour tout $\varepsilon \in (0,1)$, il existe un ensemble d'entiers Λ avec densité* $\mathrm{dens}(\Lambda) \geqslant (1-\varepsilon)^2\nu(B)$ *tel que pour tout $f \in L_2^\Lambda(\mathbb{T}, \nu)$, on a*

$$\frac{\varepsilon}{2-\varepsilon}\|f\|_{L_2(\mathbb{T})} \leqslant \|f\|_{L_2(B)} \leqslant \frac{2-\varepsilon}{\varepsilon}\|f\|_{L_2(\mathbb{T})}$$

Dans le régime où ε est proche de 1, on a le corollaire suivant :

Corollaire 2. *Soit $B \subset \mathbb{T}$ tel que $\nu(B) > 0$. Pour tout $\varepsilon \in (0,1)$, il existe un ensemble d'entiers Λ avec densité* $\mathrm{dens}(\Lambda) \geqslant \frac{\varepsilon^2}{9}\nu(B)$ *tels que pour tout $f \in L_2^\Lambda(\mathbb{T}, \nu)$, on a*

$$(1-\varepsilon)\|f\|_{L_2(\mathbb{T})} \leqslant \|f\|_{L_2(B)} \leqslant (1+\varepsilon)\|f\|_{L_2(\mathbb{T})}$$

Ceci signifie que pour tout $B \subset \mathbb{T}$ de mesure non nulle, on peut trouver un ensemble d'entiers Λ à densité non nulle tel que les deux normes $\|\cdot\|_{L_2(\mathbb{T})}$ et $\|\cdot\|_{L_2(B)}$ sont équivalentes sur $L_2^\Lambda(\mathbb{T}, \nu)$ avec un facteur d'équivalence proche de 1.

0.2 Distance de Banach-Mazur au cube

Etudier la distance entre les objets est un outil efficace pour pouvoir les identifier. Pour chaque catégorie d'objets, il faut définir la bonne notion de distance qui colle également avec la notion d'application définie sur cette catégorie. Dans le cadre des espaces de Banach, on s'intéresse à la distance de Banach-Mazur qui va mesurer à quel point deux espaces sont isomorphes. Pour avoir une idée claire, considérons un espace de Banach de dimension n comme $\mathbb{R}^n$ muni d'une norme ; mesurer la distance géométrique entre deux espaces de Banach revient à mesurer l'équivalence entre leurs normes. On sait que deux normes $|\cdot|$ et $\|\cdot\|$ sur $\mathbb{R}^n$ sont équivalentes c.à.d

$$\alpha|\cdot| \leqslant \|\cdot\| \leqslant \beta|\cdot|,$$

mais il s'agit d'estimer le facteur d'équivalence $\frac{\beta}{\alpha}$ qui peut dépendre de la dimension. Ainsi si on estime le facteur d'équivalence entre les normes de deux espaces de Banach X et Y, on aurait

$$\alpha B_X \subset B_Y \subset \beta B_X,$$

où B_X et B_Y désignent les boules unités respectives de X et Y ; la distance géométrique entre X et Y serait donc le facteur d'équivalence entre leurs normes. Pour calculer la distance de Banach-Mazur entre X et Y, il faut minimiser sur tous les $T \in GL_n(\mathbb{R})$ la distance géométrique entre X et TY. Pour commencer, on notera $\mathcal{BM}_n$ le compact de Banach-Mazur c.à.d l'ensemble des espaces de Banach de dimension n. La distance de Banach-Mazur entre deux éléments X et Y de $\mathcal{BM}_n$ est définie par :

$$d(X,Y) = \inf\{\|T\| \cdot \|T^{-1}\| \mid \ T \text{ est un isomorphisme entre X et Y } \}$$

Ainsi d va mesurer à quel point X et Y sont isomorphes ; s'ils sont isométriques alors cette distance vaut 1. Il est assez facile de voir que d est multiplicative, ce qui veut dire

$$\forall X, Y, Z \in \mathcal{BM}_n, \quad d(X,Z) \leqslant d(X,Y)d(Y,Z).$$

Ainsi $\log(d)$ est une distance sur $\mathcal{BM}_n$ quotienté par la relation d'équivalence qui consiste à identifier deux espaces isométriques. Une autre propriété importante est que d est invariante par dualité c.à.d

$$\forall X, Y \in \mathcal{BM}_n, \quad d(X,Y) = d(X^*, Y^*)$$

Estimer la distance entre deux espaces de Banach est d'une grande importance pour essayer de mieux comprendre la structure de $\mathcal{BM}_n$. La première démarche est évidemment de regarder sur des exemples et plus précisément la distance entre les espaces l_p^n dont les normes sont données par

$$\|x\|_p = \left(\sum_{j\leqslant n} |x_j|^p\right)^{\frac{1}{p}} \text{ si } 1 \leqslant p < \infty \quad \text{ et } \quad \|x\|_\infty = \max_{j\leqslant n} |x_j|,$$

où $x = (x_j)_{j\leqslant n} \in \mathbb{R}^n$. Il est assez facile de voir que $d(l_1^n, l_2^n) = d(l_\infty^n, l_2^n) = \sqrt{n}$ et plus généralement que $d(l_p^n, l_q^n) = n^{\frac{1}{p}-\frac{1}{q}}$ si $1 \leqslant p \leqslant q \leqslant 2$ ou $2 \leqslant p \leqslant q \leqslant \infty$. Dans le cas où $1 \leqslant p < 2 < q \leqslant \infty$, il a été démontré dans [37] que

$$\frac{1}{\sqrt{2}} \max\left(n^{\frac{1}{p}-\frac{1}{2}}, n^{\frac{1}{2}-\frac{1}{q}}\right) \leqslant d(l_p^n, l_q^n) \leqslant (1+\sqrt{2}) \max\left(n^{\frac{1}{p}-\frac{1}{2}}, n^{\frac{1}{2}-\frac{1}{q}}\right),$$

et en particulier que $d(l_1^n, l_\infty^n)$ est de l'ordre de $\sqrt{n}$. Mis à part quelques exemples et cas particuliers, estimer la distance de Banach-Mazur en toute généralité s'avère être un problème difficile. Le résultat de John [40] qu'on a énoncé dans le Théorème 0.2 permet d'affirmer que la distance d'un élément quelconque de $\mathcal{BM}_n$ à l_2^n est au plus $\sqrt{n}$.

Si on note

$$R_2^n = \max \left\{ d(X, l_2^n) / \ X \in \mathcal{BM}_n \right\},$$

alors le théorème de John dit que $R_2^n \leqslant \sqrt{n}$ et comme $d(l_1^n, l_2^n) = \sqrt{n}$ alors $R_2^n = \sqrt{n}$. Par multiplicativité de la distance, on a que $d(X, Y) \leqslant n$ pour tout $X, Y \in \mathcal{BM}_n$ et ainsi que le diamètre de $\mathcal{BM}_n$ est inférieur ou égal à n. Gluskin [34] a montré que n est le bon ordre de grandeur pour le diamètre de $\mathcal{BM}_n$ en construisant deux espaces X et Y tels que $d(X, Y) \geqslant cn$.

Comme le diamètre de $\mathcal{BM}_n$ est de l'ordre de n et que $R_2^n = \sqrt{n}$ alors l_2^n est un centre du compact de Banach-Mazur $\mathcal{BM}_n$.

0.2.1 Majoration de la distance de Banach-Mazur au cube

Dans cette thèse, on s'est intéressé à la distance de Banach-Mazur à l_∞^n, dont la boule unité est un cube de dimension n. On a déjà vu que la distance des espaces l_p^n à l_∞^n, et plus surprenant de l_1^n à l_∞^n, ne dépasse pas $\sqrt{n}$ asymptotiquement. Une question naturelle serait de se demander si l_∞^n est également un centre de $\mathcal{BM}_n$ et que vaut la distance d'un espace de Banach quelconque à l_∞^n. Par une construction similaire à celle de Gluskin, Szarek [80] a montré l'existence d'un espace de Banach X tels que $d(X, l_\infty^n) \geqslant c\sqrt{n}\log(n)$, ce qui signifie que l_∞^n n'est pas un centre de $\mathcal{BM}_n$. Notons

$$R_\infty^n = \max \left\{ d(X, l_\infty^n) / \ X \in \mathcal{BM}_n \right\} \quad \text{et} \quad R_1^n = \max \left\{ d(X, l_1^n) / \ X \in \mathcal{BM}_n \right\}.$$

Par dualité, on a que $R_\infty^n = R_1^n$ et par le résultat de Szarek [80], on a que $R_1^n \geqslant c\sqrt{n}\log(n)$. Le problème de donner une majoration de R_∞^n a d'abord été abordé par Bourgain-Szarek [18] qui ont démontré que $R_\infty^n = o(n)$. Szarek-Talagrand [82] puis Giannopoulos [31] ont amélioré cette estimation pour obtenir $cn^{\frac{7}{8}}$ et $cn^{\frac{5}{6}}$ respectivement. Ces preuves reposent sur une factorisation du type Dvoretzky-Rogers dont on discutera dans la partie suivante. Notons que Taschuk [84] a démontré une estimation de R_∞^n pour les petites dimensions ; précisément

$$R_\infty^n \leqslant \sqrt{n^2 - 2n + 2 + \frac{2}{\sqrt{n+2} - 1}},$$

ce qui n'est pas satisfaisant pour les grandes dimensions puisque c'est de l'ordre de n.

Nous proposons une nouvelle preuve de ce qui est appelé "proportional Dvoretzky-Rogers factorization" et qui constitue le coeur des méthodes précédentes. Ceci sera détaillé par la suite, mais ce résultat nous permet d'améliorer les constantes intervenant dans l'estimation finale de R_∞^n tout en retrouvant le même ordre asymptotique.

Théorème 5. *Soit X un espace de Banach de dimension n. Alors*

$$d(X, l_\infty^n) \leqslant 2^{\frac{4}{3}}\sqrt{n} \cdot d(X, l_2^n)^{\frac{2}{3}}.$$

On montre également que $R_\infty^n \leqslant (2n)^{\frac{5}{6}}$.

Notons également que cette estimation améliore celle de Taschuk à partir de la dimension 22.

0.2.2 "Proportional Dvoretzky-Rogers factorization" : cadre symétrique

Le lemme de Dvoretzky-Rogers [28] affirme que si X est un espace de Banach de dimension n alors il existe $x_1, ..., x_k \in X$ avec $k = \sqrt{n}$ tels que pour toute suite de scalaires $(a_j)_{j\leqslant k}$ on a

$$\max_{j\leqslant k} |a_j| \leqslant \left\| \sum_{j\leqslant k} a_j x_j \right\|_X \leqslant c \left(\sum_{j\leqslant k} a_j^2 \right)^{\frac{1}{2}},$$

où c est une constante universelle. Bourgain-Szarek [18] ont montré que ce résultat reste vrai pour k proportionnel à n, d'où l'appellation "the proportional Dvoretzky-Rogers factorization".

Théorème 0.10 (Proportional Dvoretzky-Rogers factorization)**.** *Soit X un espace de Banach de dimension n. Pour tout $\varepsilon \in (0,1)$, il existe $x_1, ..., x_k \in X$ avec $k \geqslant (1-\varepsilon)n$ tels que pour toute suite de scalaires $(a_j)_{j\leqslant k}$, on a*

$$\max_{j\leqslant k} |a_j| \leqslant \left\| \sum_{j\leqslant k} a_j x_j \right\|_X \leqslant c(\varepsilon) \left(\sum_{j\leqslant k} a_j^2 \right)^{\frac{1}{2}},$$

où $c(\varepsilon)$ est une constante qui dépend de ε. De manière équivalente, l'identité $i_{2,\infty} : l_2^k \longrightarrow l_\infty^k$ s'écrit $i_{2,\infty} = \alpha \circ \beta$ où $\beta : l_2^k \longrightarrow X, \alpha : X \longrightarrow l_\infty^k$ et $\|\alpha\| \cdot \|\beta\| \leqslant c(\varepsilon)$.

Trouver la bonne dépendance en ε est un problème important et reste à présent un problème ouvert. Szarek [80] a montré qu'on ne peut espérer une dépendance meilleure que $c\varepsilon^{-\frac{1}{10}}$. Szarek-Talagrand [82] ont montré le Théorème 0.10 avec $c(\varepsilon) = c\varepsilon^{-2}$ alors que Giannopoulos a amélioré ceci pour obtenir $c\varepsilon^{-\frac{3}{2}}$ et $c\varepsilon^{-1}$ dans [31] et [32] respectivement. Dans tous ces résultats, une factorisation pour l'identité $i_{1,2}$ a été donnée, la factorisation de $i_{2,\infty}$ découlant par dualité. Les preuves utilisaient des arguments géométriques, dont le lemme de Dvoretzky-Rogers, et des résultats combinatoires assez techniques ainsi que le théorème de factorisation de Grothendieck.

Nous proposons une approche plus élémentaire basée sur le principe d'invertibilité restreinte normalisé qu'on a montré dans le Théorème 1. Ceci nous permet de simplifier considérablement la preuve et montrer ce qui suit :

Théorème 6. *Soit X un espace de Banach de dimension n. Pour tout $\varepsilon \in (0,1)$, il existe $x_1, ..., x_k \in X$ avec $k \geqslant (1-\varepsilon)^2 n$ tel que pour toute famille de scalaires $(a_j)_{j\leqslant k}$*

$$\varepsilon \left(\sum_{j\leqslant k} a_j^2 \right)^{\frac{1}{2}} \leqslant \left\| \sum_{j\leqslant k} a_j x_j \right\|_X \leqslant \sum_{j\leqslant k} |a_j|$$

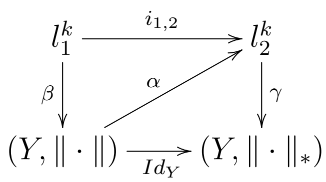
De manière équivalente, l'identité $i_{1,2} : l_1^k \longrightarrow l_2^k$ s'écrit $i_{1,2} = \alpha \circ \beta$, où $\beta : l_1^k \longrightarrow X$, $\alpha : X \longrightarrow l_2^k$ et $\|\alpha\| \cdot \|\beta\| \leqslant \varepsilon^{-1}$.

Comme application directe de cette factorisation, on a le corollaire suivant :

Corollaire 3. *Soit X un espace de Banach de dimension n. Pour tout $\varepsilon \in (0,1)$, il existe Y un sous-espace de X de dimension $k \geqslant [(1-\varepsilon)^2 n]$ tels que $d(Y, l_1^k) \leqslant \frac{\sqrt{n}}{\varepsilon}$.*

Se servant du Théorème 3, on obtient un résultat de factorisation plus général :

Théorème 7. *Soit $X = (\mathbb{R}^n, \|\cdot\|)$ où $\|\cdot\|$ est une norme sur $\mathbb{R}^n$ telle que B_2^n est l'ellipsoide de volume minimal contenant B_X. Pour tout $\varepsilon \in (0,1)$, il existe $Y \subset \mathbb{R}^n$ un sous-espace de dimension $k \geqslant (1-\varepsilon)^2 n$ tel que le diagramme suivant commute*



où $\|\cdot\|_*$ est la norme duale de $\|\cdot\|$. Cela veut dire que $i_{1,2} = \alpha \circ \beta$ et $Id_Y = \gamma \circ \alpha$ et de plus, $\|\beta\| \leqslant 1$, $\|\alpha\| \leqslant \frac{2-\varepsilon}{\varepsilon}$ et $\|\gamma\| \leqslant \frac{2-\varepsilon}{\varepsilon}$.

0.2.3 "Proportional Dvoretzky-Rogers factorization" : cadre non symétrique

Les résultats précédents peuvent être énoncés en termes de corps convexes symétriques puisqu'on a déjà vu que la donnée d'une norme sur $\mathbb{R}^n$ est équivalente à la donnée d'un corps convexe symétrique. Dans [50], Litvak et Tomczak-Jaegermann ont montré une version non symétrique du Théorème 0.10. Précisément, l'énoncé est le suivant :

Théorème 0.11 (Litvak-Tomczak-Jaegermann)**.** *Soit $K \subset \mathbb{R}^n$ un corps convexe, tel que B_2^n est l'ellipsoïde de volume minimal contenant K. Soient $\varepsilon \in (0,1)$ et $k = [(1-\varepsilon)n]$. Il existe $y_1, y_2, ..., y_k$ dans K, et P une projection orthogonale dans $\mathbb{R}^n$ de rang supérieur ou égal à k tels que pour toute famille de scalaires $(a_j)_{j\leqslant k}$ on a*

$$c\varepsilon^3 \left(\sum_{j=1}^{k} |a_j|^2 \right)^{\frac{1}{2}} \leqslant \left\| \sum_{j=1}^{k} a_j P y_j \right\|_{PK} \leqslant \frac{6}{\varepsilon} \sum_{j=1}^{k} |a_j|,$$

où $c > 0$ est une constante universelle.

Notre but est, comme dans le cas symétrique, de donner une nouvelle preuve de ce résultat. En introduisant ce nouvel ingrédient fournit par le Théorème 1, nous améliorons les dépendances en ε dans le résultat précédent et nous obtenons :

Théorème 8. *Soit $K \subset \mathbb{R}^n$ un corps convexe tel que B_2^n est l'ellipsoïde de volume minimal contenant K. Soient $\varepsilon \in (0,1)$ et $k = [(1-\varepsilon)n]$. Il existe $y_1, y_2, ..., y_k$ dans K, et P une projection orthogonale dans $\mathbb{R}^n$ de rang supérieur ou égal à k tels que pour toute famille de scalaires $(a_j)_{j\leqslant k}$ on a*

$$\frac{\varepsilon^2}{16} \left(\sum_{j=1}^{k} |a_j|^2 \right)^{\frac{1}{2}} \leqslant \left\| \sum_{j=1}^{k} a_j P x_j \right\|_{PK} \leqslant \frac{4}{\varepsilon} \sum_{j=1}^{k} |a_j|$$

Comme application directe de ce résultat, on déduit :

Corollaire 4. *Soit $K \subset \mathbb{R}^n$ un corps convexe tels que B_2^n est l'ellipsoïde de volume minimal contenant K. Pour tout $\varepsilon \in (0,1)$, il existe P une projection orthogonale de $\mathbb{R}^n$ de rang $k \geqslant [(1-\varepsilon)n]$ tel que*

$$\frac{\varepsilon}{4} B_1^k \subset PK \subset \frac{16}{\varepsilon^2} B_2^k.$$

En plus, $d(PK, B_1^k) \leqslant \frac{64\sqrt{n}}{\varepsilon^3}$.

Par dualité, il existe un sous-espace $E \subset \mathbb{R}^n$ *de dimension* $k \geqslant [(1-\varepsilon)n]$ *tels que*

$$\frac{\varepsilon^2}{16} B_2^k \subset K \cap E \subset \frac{4}{\varepsilon} B_\infty^k.$$

En plus, $d(K \cap E, B_\infty^k) \leqslant \frac{64\sqrt{n}}{\varepsilon^3}$.

Il est intéressant de noter que la dépendance en la dimension dans le résultat précédent est la même que dans le cadre symétrique qu'on a vu dans le Corollaire 3.

0.3 Covariance de matrices aléatoires

Les matrices aléatoires constituent un sujet très vaste dont les applications touchent à de nombreux domaines des Mathématiques. Dans cette thèse, nous n'abordons que le problème d'estimation de la matrice de covariance et plus précisément l'extension au cadre matriciel de certains résultats vectoriels.

Si X est un vecteur aléatoire de $\mathbb{R}^n$, il s'agit d'approcher la matrice de covariance de X qui est donnée par $\mathbb{E}XX^t$. Par la loi des grands nombres, on sait qu'en prenant beaucoup de copies indépendantes de X, la moyenne convergera vers la matrice de covariance de X. Plus précisément, si $(X_j)_{j\leqslant N}$ sont des copies indépendantes de X alors

$$\frac{1}{N}\sum_{j\leqslant N} X_j X_j^t \underset{N\to\infty}{\longrightarrow} \mathbb{E}XX^t \quad \text{p.s}$$

On s'intéresse au problème de quantifier cette convergence c.à.d déterminer le nombre minimal de copies nécessaires pour bien approcher la matrice de covariance de X. Ce problème se formule donc comme suit :

Pour $\varepsilon > 0$, trouver le nombre minimal $N = N(n, \varepsilon)$ de copies indépendantes de X tels que

$$\left\| \frac{1}{N}\sum_{j\leqslant N} X_j X_j^t - \mathbb{E}XX^t \right\| \leqslant \varepsilon,$$

avec grande probabilité ou même en espérance.

Ceci rappelle le problème d'approximation de l'identité étudié par Batson-Spielman-Srivastava [13]. En effet, soit $Id = \sum_{j\leqslant m} x_j x_j^t$ une décomposition de l'identité sur $\mathbb{R}^n$ et $\varepsilon > 0$. On définit

X un vecteur aléatoire sur $\mathbb{R}^n$ par

$$X = \frac{\sqrt{n}}{\|x_j\|_2} x_j \quad \text{avec probabilité } p_j = \frac{\|x_j\|_2^2}{n}.$$

Il est facile de voir que $\sum_{j \leqslant m} p_j = 1$ et que

$$\mathbb{E}XX^t = \sum_{j \leqslant} p_j \frac{n}{\|x_j\|_2^2} x_j x_j^t = Id.$$

Supposons qu'on sache estimer la matrice de covariance de X c.à.d en prenant $X_1, ..., X_N$ des copies indépendantes de X on a

$$\left\| \frac{1}{N} \sum_{j \leqslant N} X_j X_j^t - Id \right\| \leqslant \varepsilon,$$

avec probabilité strictement positive. Cela veut dire qu'il existe $\sigma \subset \{1, ..., m\}$ de taille N tel que

$$\left\| \sum_{j \leqslant m} s_j x_j x_j^t - Id \right\| \leqslant \varepsilon \quad \text{et} \quad \{j, s_j \neq 0\} = \sigma.$$

Si N était de l'ordre de n, on retrouverait le Théorème 0.1. Cependant, en suivant cette démarche, Rudelson [67] obtient N de l'ordre de $n \log(n)$ ce qui est un résultat optimal. Néanmoins, ceci montre le lien entre le problème d'approximation d'une décomposition de l'identité et celui de l'approximation de la matrice de covariance d'un vecteur aléatoire.

Notons qu'il est assez facile d'obtenir l'approximation de la matrice de covariance d'un vecteur gaussien avec un nombre de copies indépendantes proportionnel à la dimension ; en effet, les vecteurs gaussiens bénéficient de propriétés de concentration avec vitesse exponentielle qui permettent d'effectuer l'argument standard de réseau sur la sphère unité pour pouvoir estimer

$$\sup_{x \in S^{n-1}} \left| \sum_{j \leqslant N} \langle G_j, x \rangle^2 - 1 \right|,$$

où G_j désignent des vecteurs gaussiens standards indépendants. Cette stratégie n'est plus valable en toute généralité. Après une suite de travaux consacrés à ce problème ([4], [5],[3]), Adamczack et al. ont réussi à démontrer qu'un nombre proportionnel à la dimension suffit pour approcher la matrice de covariance d'un vecteur aléatoire bénéficiant de propriétés de concentration avec vitesse sous-exponentielle. On réfère à [4] pour des énoncés plus précis et plus de références sur ce problème. Dans cette thèse, on s'est intéressé à un résultat dù à Srivastava-Vershynin [79] sur le problème d'approximation de la matrice de covariance.

Théorème 0.12 (Srivastava-Vershynin). *Considérons X_i des vecteurs aléatoires indépendants isotropes dans $\mathbb{R}^n$. Supposons que X_i satisfait une hypothèse de régularité (SR) : il existe $C, \eta > 0$ telle que pour toute projection orthogonale P de $\mathbb{R}^n$,*

$$\mathbb{P}\left\{\|PX_i\|_2^2 > t\right\} \leqslant Ct^{-1-\eta} \quad \forall t > C\mathrm{rang}(P). \tag{5}$$

Alors, pour tout $\varepsilon \in (0,1)$ et pour

$$N \geqslant C\,\varepsilon^{-2-2/\eta} \cdot n$$

on a

$$\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N} X_i X_i^t - Id\right\| \leqslant \varepsilon. \tag{6}$$

La preuve de ce théorème consiste à rendre aléatoire la méthode de Batson-Spielman-Srivastava [13]. Ce résultat a l'avantage de couvrir une grande classe de distributions ; les vecteurs log-concave ainsi que les vecteurs aléatoires ayant des entrées indépendantes avec seulement des moments d'ordre plus grand que 2 vérifient (5). Cependant, l'estimation de la matrice de covariance est seulement en espérance.

Notre but est d'étendre ce résultat à un cadre matriciel, dans le sens de remplacer le vecteur aléatoire X par une matrice aléatoire de taille quelconque. Notons que plusieurs travaux ont été consacrés à étendre des résultats connus pour des variables/vecteurs aléatoires à un cadre matriciel (voir par exemple [6], [54] et [88]).

0.3.1 Estimation de la covariance de matrices aléatoires

Comme nous l'avons déjà dit, nous nous intéressons à étendre le Théorème 0.12 à un cadre matriciel c.à.d étant donné A une matrice aléatoire de taille $n \times m$, il s'agit d'approcher $\mathbb{E}AA^t$. Le cadre intéressant est quand les colonnes de A ne sont pas indépendantes ; sinon on écrit $AA^t = \sum_{j\leqslant m} C_j C_j^t$, $(C_j)_{j\leqslant m}$ étant les colonnes de A, et si chaque colonne vérifie (5) alors le Théorème 0.12 s'applique pour chaque colonne et ainsi par l'inégalité triangulaire l'approximation de $\mathbb{E}AA^t$ en découlerait. Notre problème peut être formulé d'une manière plus générale :

Soit B une matrice aléatoire symétrique semi-définie positive de taille $n \times n$ vérifiant certaines hypothèses de régularité. Trouver le nombre minimal N de copies indépendantes de B

tels que $\mathbb{E}\left\|\frac{1}{N}\sum_{i\leqslant N} B_i - \mathbb{E}B\right\|$ soit petit.

Si on suppose que $\mathbb{E}B = Id$ et que $\|B\| \leqslant n$ presque surement, alors $cn\log(n)$ copies indépendantes de B suffisent pour l'approximation. En effet, prenons $B_1, ..., B_N, B'_1, ..., B'_N$ des copies indépendantes de B et $\varepsilon_1, ..., \varepsilon_N$ des bernoullis ± 1 indépendantes et écrivons

$$
\begin{aligned}
\alpha := \mathbb{E}\left\|\frac{1}{N}\sum_{i\leqslant N} B_i - Id\right\| &= \mathbb{E}\left\|\frac{1}{N}\sum_{i\leqslant N}(B_i - \mathbb{E}B'_i)\right\| \\
&\leqslant \mathbb{E}\left\|\frac{1}{N}\sum_{i\leqslant N} B_i - B'_i\right\| \quad \text{par Jensen} \\
&\leqslant 2\mathbb{E}\left\|\frac{1}{N}\sum_{i\leqslant N}\varepsilon_i B_i\right\| \quad \text{par symétrisation} \\
&\leqslant \frac{2}{N}\left(\mathbb{E}\left\|\sum_{i\leqslant N}\varepsilon_i B_i\right\|^p\right)^{\frac{1}{p}} \quad \text{pour tout } p \geqslant 1
\end{aligned}
$$

En utilisant l'inégalité de Khintchine non commutative ([52], [53], voir également [57]), on a

$$
\left(\mathbb{E}_\varepsilon\left\|\sum_{i\leqslant N}\varepsilon_i B_i\right\|^p\right)^{\frac{1}{p}} \leqslant (c\sqrt{\log(n)} + \sqrt{p})\left\|\sum_{i\leqslant N} B_i^2\right\|^{\frac{1}{2}}
$$

Comme les B_i sont semi-définies positives alors

$$
\left\|\sum_{i\leqslant N} B_i^2\right\|^{\frac{1}{2}} \leqslant \max_{i\leqslant N}\|B_i\|^{\frac{1}{2}} \cdot \left\|\sum_{i\leqslant N} B_i\right\|^{\frac{1}{2}}
$$

Ainsi si $\|B\| \leqslant n$ presque surement, on a d'après ce qui précède

$$
\begin{aligned}
\alpha &\leqslant \frac{c\sqrt{n\log(n)}}{N} \cdot \mathbb{E}\left\|\sum_{i\leqslant N} B_i\right\|^{\frac{1}{2}} \\
&\leqslant \sqrt{\frac{n\log(n)}{N}} \cdot \left[\mathbb{E}\left\|\sum_{i\leqslant N}\frac{1}{N}B_i\right\|\right]^{\frac{1}{2}} \\
&\leqslant c\sqrt{\frac{n\log(n)}{N}} \cdot \left(\sqrt{\alpha} + 1\right)
\end{aligned}
$$

Par un simple calcul, on déduit que

$$\alpha = \mathbb{E}\left\| \frac{1}{N} \sum_{i \leqslant N} B_i - Id \right\| \leqslant c\sqrt{\frac{n\log(n)}{N}}$$

et ainsi il suffirait de prendre N de l'ordre de $cn\log(n)$ pour rendre α petit.

Sans hypothèses de régularité, on ne peut pas dire mieux que ça ; en effet, en prenant B uniformément distribuée sur $\{ne_ie_i^t\}_{i\leqslant n}$, $(e_i)_{i\leqslant n}$ étant la base canonique de $\mathbb{R}^n$, on peut voir qu'on a besoin d'au moins $cn\log(n)$ copies.

De Carli Silva-Harvey-Sato [26] ont pu étendre le résultat de Batson-Spielman-Srivastava [13] à un cadre matriciel. En rendant alétaoire la méthode de De Carli Silva-Harvey-Sato, on obtient le résultat suivant :

Théorème 9. *Soit B une matrice aléatoire symétrique semi-définie positive de taille $n \times n$. Supposons que $\mathbb{E}B = Id$ et que B satisfait une hypothèse de régularité (MSR) : il existe $c, \eta > 0$ telle que pour toute projection orthogonale P de $\mathbb{R}^n$,*

$$\mathbb{P}\left\{\|PBP\| > t\right\} \leqslant ct^{-1-\eta} \quad \forall t > c\mathrm{rang}(P). \tag{7}$$

Alors, pour tout $\varepsilon \in (0,1)$ et pour

$$N \geqslant c\,\varepsilon^{-2-2/\eta} \cdot n$$

on a

$$\mathbb{E}\left\| \frac{1}{N} \sum_{i=1}^{N} B_i - Id \right\| \leqslant \varepsilon, \tag{8}$$

où $B_1, ..., B_N$ sont des copies indépendantes de B.

Ceci généralise le Théorème 0.12 ; en effet, soit X est un vecteur aléatoire isotrope de $\mathbb{R}^n$ qui vérifie (5). Posons $B = XX^t$, alors $\mathbb{E}B = Id$ et comme $\|PBP\| = \|PX\|_2^2$ alors B vérifie (7). Ainsi en appliquant le Théorème 9 à B, on retrouve (6).

Afin de démontrer le Théorème 9, nous estimons la plus petite et la plus grande valeur propre de la somme de matrices aléatoires vérifiant (MSR).

Dans [9], Bai-Yin ont démontré que si X est un vecteur aléatoire de $\mathbb{R}^n$ ayant des entrées i.i.d centrées de variance 1 et dont le moment d'ordre 4 est fini, alors pour $N = N(n)$ tels que

$h = \lim_{n\to\infty} \frac{n}{N} \in (0,1)$ on a

$$\lambda_{min}\left(\frac{1}{N}\sum_{j\leqslant N} X_j X_j^t\right) \underset{n\to\infty}{\longrightarrow} (1-\sqrt{h})^2 \quad p.s$$

et

$$\lambda_{max}\left(\frac{1}{N}\sum_{j\leqslant N} X_j X_j^t\right) \underset{n\to\infty}{\longrightarrow} (1+\sqrt{h})^2 \quad p.s$$

Les estimations de la plus petite et la plus grande valeur propre de la somme de matrices aléatoires vérifiant (MSR), données par le Théorème 9, peuvent être interprétées comme une version matricielle non asymptotique du résultat de Bai-Yin.

Nous donnons des exemples d'applications de nos résultats. Le cas des matrices log-concave en est un et sera détaillé par la suite.

0.3.2 Matrices log-concave

Une mesure de probabilité μ sur $\mathbb{R}^n$ est dite log-concave si pour tout $0 < t < 1$ et pour tout $A, B \subset \mathbb{R}^n$ compacts de mesures non nulles, on a

$$\mu\left((1-t)A + tB\right) \geqslant \mu(A)^{1-t}\mu(B)^t.$$

Borell ([17],[16]) a donné une caractérisation des mesures log-concave : μ est une mesure log-concave sur $\mathbb{R}^n$ si et seulement si la densité f de μ par rapport à la mesure de Lebesgue est log-concave i.e $\log(f)$ est une fonction concave. Finalement, un vecteur aléatoire de $\mathbb{R}^n$ est dit log-concave si sa distribution est log-concave.

Les mesures log-concave jouent un rôle très important dans la géométrie des convexes ; en effet, si K est un corps convexe alors $\mathbf{1}_K$ est une mesure log-concave. D'autre part, étant donné une mesure log-concave, Ball [10] associe à cette mesure un corps convexe ; d'où leur appellation les corps de Ball. Beaucoup de travaux concernant les mesures log-concave ont été réalisés ces dernières années ; un des résultats importants est celui de Paouris ([58], [59], [60]) qui montre des inégalités de grandes déviations et des estimations de petites boules pour les vecteurs log-concave. Pour faire le lien avec la partie précédente, un des problèmes importants est celui d'estimer la matrice de covariance d'un vecteur log-concave ; ceci a été résolu par Adamczack et al. [4]. Nous nous sommes intéressés à la notion de log-concavité dans le cadre matriciel. Vu qu'une matrice de taille $n \times m$ peut être vue comme un vecteur de $\mathbb{R}^{nm}$, ainsi on peut définir d'une manière naturelle la notion de matrices log-concave ; précisément, une matrice aléatoire

A de taille $n \times m$ est log-concave si sa distribution est une mesure log-concave sur $\mathbb{R}^{nm}$. Il reste donc à définir la notion d'isotropie.

Définition 0.1. *Soit A une matrice aléatoire de taille $n \times m$ ayant $(C_i)_{i \leqslant m}$ comme colonnes. On dira que A est une matrice log-concave isotrope si $A'^t = \sqrt{m}(C_1^t, .., C_m^t)$ est un vecteur log-concave isotrope de $\mathbb{R}^{nm}$.*

Ainsi lorsque A est isotrope, on a que $\mathbb{E}AA^t = Id$. Il est assez facile de déduire des propriétés de concentrations de ces matrices à partir des résultats de Paouris. Nous utilisons cependant un résultat plus raffiné dû à Guédon-Milman [36], qui affirme qu'avec grande probabilité, un vecteur log-concave isotrope de $\mathbb{R}^n$ est situé dans une fine couronne dont le rayon est presque $\sqrt{n}$; ceci est connu sous le nom de thin-shell. Ainsi les matrices log-concave vérifient ce qui suit :

Proposition 5. *Soit A une matrice log-concave isotrope de taille $n \times m$. Si $B = AA^t$, alors pour toute projection orthogonale de $\mathbb{R}^n$, on a*

$$\mathbb{P}\left\{|\mathrm{Tr}(PB) - \mathrm{rang}(P)| \geqslant t.\mathrm{rang}(P)\right\} \leqslant C \exp\left(-ct^3\sqrt{m.\mathrm{rang}(P)}\right) \quad \forall t \leqslant 1. \tag{9}$$

A partir de ça, il est assez facile de voir qu'une matrice log-concave isotrope vérifie (MSR) et donc les résultats de la section précédente s'y appliquent.

Les propriétés de concentration données par la Proposition 5 sont suffisantes pour qu'on puisse obtenir des résultats avec grande probabilité et non seulement en espérance. On a cependant besoin que la matrice soit suffisamment rectangulaire. Plus précisément, on montre ce qui suit :

Théorème 10. *Soit A une matrice log-concave isotrope de taille $n \times m$. Pour tout $\varepsilon \in (0,1)$, si $m \geqslant \frac{C}{\varepsilon^6}\left[\log(CnN)\right]^2$, en prenant*

$$N \geqslant \frac{96n}{\varepsilon^2}$$

copies indépendantes $A_1, ..., A_N$ de A, alors avec probabilité $\geqslant 1 - \exp(-c\varepsilon^3\sqrt{m})$ on a

$$\left\| \frac{1}{N}\sum_{i=1}^N A_i A_i^t - Id \right\| \leqslant \varepsilon$$

Notons que le nombre de copies indépendantes nécéssaire dans le résultat précédent est optimal comme on peut voir en prenant des matrices gaussiennes. L'ensemble des matrices log-concave isotropes qu'on a défini contient une large classe d'exemples. Ainsi on a ce qui suit :

Proposition 6. *Soit A une matrice aléatoire de taille $n \times m$ dont la densité par rapport à la mesure de Lebesgue est donnée par*

$$G(A) = \exp\left(-f(s_1(A), ..., s_k(A))\right),$$

où f est une fonction convexe absolument symétrique, proprement normalisée et $k = \min(n, m)$. Supposons que $m \geqslant \frac{C}{\varepsilon^6}\left[\log(\frac{Cn}{\varepsilon})\right]^2$ et $n \geqslant \frac{C}{\varepsilon^6}\left[\log(\frac{Cm}{\varepsilon})\right]^2$, en prenant $N = \frac{96\max(n,m)}{\varepsilon^2}$ alors avec probabilité $\geqslant 1 - \exp(-c\varepsilon^3\sqrt{k})$ on a

$$1 - \varepsilon \leqslant \lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^{N} A_i A_i^t\right) \leqslant \lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} A_i A_i^t\right) \leqslant 1 + \varepsilon$$

et

$$(1-\varepsilon)\frac{n}{m} \leqslant \lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^{N} A_i^t A_i\right) \leqslant \lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} A_i^t A_i\right) \leqslant (1+\varepsilon)\frac{n}{m}$$

Récemment, une nouvelle preuve du résultat de Paouris a été donnée dans [2]. Dans [1], il a été démontré que la méthode de [2] s'étend au cadre des mesures convexes. Ainsi, des propriétés de concentration pour les vecteurs $(-\frac{1}{r})$-concaves isotropes ont été établies. A partir de ces propriétés, on peut obtenir des analogues des résultats précédents dans le cadre des matrices $(-\frac{1}{r})$-concaves isotropes.

0.4 Références bibliographiques

Cette thèse regroupe trois papiers qui ont tous été soumis à des revues internationales. Ainsi, le contenu des chapitres est essentiellement le contenu de ces articles. Notons qu'il y a quelques résultats supplémentaires qui ont été rajoutés. Nos papiers peuvent être trouvés sur Arxiv :

1. "Restricted invertibility and the banach-mazur distance to the cube", Available at arXiv :1206.0654.
2. "A note on column subset selection", Available at arXiv :1212.0976.
3. "Estimating the covariance of random matrices", Available at arXiv :1301.6607.

Les résultats des chapitres 1 et 3 sont essentiellement contenus dans le premier papier. Les résultats du chapitre 2 font partie du deuxième papier, et finalement les résultats des chapitres 4 et 5 sont contenus dans le troisième papier.

Introduction

This thesis lies in the field of functional analysis. We are interested in the local theory of Banach spaces i.e their finite-dimensional structures. A Banach space of dimension n can be considered as $\mathbb{R}^n$ equipped with a norm. As in each class of objects, we study the tools that go with it (applications, distance, etc. ..). In our framework, the applications are linear operators or simply matrices. That's why we added another area of interest which is matrix theory. The probabilistic methods and the concentration of measure phenomena are a powerful tool to solve problems arising in the theory of Banach spaces. In this context, we are interested in the study of random matrices. Our results are mainly situated in three topics : Column subset selection in a matrix, Banach-Mazur distance to the cube and estimating the covariance of random matrices. The guiding thread of these three topics, is a method of approximation of the identity invented by Batson-Spielman-Srivastava [13].

Theorem 0.1 (Batson-Spielman-Srivastava)**.** *Let $\varepsilon \in (0,1)$ and $m, n \in \mathbb{N}$. For any $x_1, \dots, x_m \in \mathbb{R}^n$ there exists $s_1, \dots, s_m \in [0, \infty)$ such that*

$$\left|\left\{i \in \{1, \dots, m\} : s_i \neq 0\right\}\right| \leqslant \left\lceil \frac{n}{\varepsilon^2} \right\rceil, \tag{1}$$

and for any $y \in \mathbb{R}^n$ we have

$$(1-\varepsilon)^2 \sum_{i=1}^{m} \langle x_i, y\rangle^2 \leqslant \sum_{i=1}^{m} s_i \langle x_i, y\rangle^2 \leqslant (1+\varepsilon)^2 \sum_{i=1}^{m} \langle x_i, y\rangle^2. \tag{2}$$

This theorem reveals that in order to approximate a matrix of the form UU^t, U being an

$n \times m$ matrix, the number of columns required is of order n. However, the selected columns must be multiplied by some weights. At first, this result was destined for the field of computer science where data reduction is very important. It completes many studies, especially due to Spielman and Teng ([75],[76]) who had already obtained similar results with some parasite log-terms. The key point in this result is certainly the method which produces a deterministic algorithm to perform the extraction of the columns. It is based on the study of the evolution of the eigenvalues of a positive matrix when perturbated by a rank 1 matrix. Later, we will be focusing on the method rather than on the result of Batson-Spielman-Srivastava [13] itself.

We can begin by noting that the first opening to the local theory of Banach spaces was the application, due to Srivastava [77], of Theorem 0.1 to the problem of approximating a convex body by a neighboring one having less contact points with the ellipsoid of maximal volume contained in it. Recall that a convex body K is a convex compact set with non-empty interior ; it is symmetric if $K = -K$. There is a correspondence between the set of symmetric convex bodies in $\mathbb{R}^n$ and n-dimensional normed spaces. Indeed, it is enough to associate to every symmetric convex body K its gauge $p_k(x) = \inf\{\lambda > 0/\ x \in \lambda K\}$ which is a norm on $\mathbb{R}^n$. Reciprocally, given a norm $\|\cdot\|$ on $\mathbb{R}^n$, one can see that $\{x \in \mathbb{R}^n/\ \|x\| \leqslant 1\}$ is a symmetric convex body. A fundamental result is due to John[40] :

Theorem 0.2 (John)**.** *Let K be a convex body in $\mathbb{R}^n$. Then, B_2^n is the ellipsoid of maximal volume contained in K if and only if $B_2^n \subset K$ and there exist $x_1, ..., x_m$ contact points of K with B_2^n and positive scalars $c_1, ..., c_m > 0$ such that*

$$Id = \sum_{j\leqslant m} c_j x_j x_j^t \qquad et \qquad \sum_{j\leqslant m} c_j x_j = 0$$

When B_2^n is the maximal volume ellipsoid contained in K, we will say that K is in John's position. Note that the previous result is still valid when B_2^n is the minimal volume ellipsoid containing K. Later, multiple generalizations of this result were given ([12],[30],[35],[49]). The contact points play an important role since they characterize somehow, the associated convex body. The number of contact points intervening in John's decomposition is smaller than $\frac{n(n+1)}{2}$, but nothing better can be said without further assumptions. An interesting problem is to reduce the number of contact points ; finding a convex body L near K i.e $L \subset K \subset \alpha L$, where α is a constant, such that L has less contact points. Approximating the identity decomposition associated to a convex body, and following Rudelson's procedure [65], Srivastava gave a first application of Theorem 0.1 to the problem of reducing the contact points number, thus improving Rudelson's result [65]. From here, the door was wide open to other applications of this result ([33],[72],[71]). In this thesis, the Batson-Spielman-Srivastava method will be pertinent since it intervenes in the three addressed themes. We will start by introducing each of the problems that we treated :

0.1 Restricted invertibility and column subset selection

Let U be an $n \times m$ matrix. We see U as a linear operator from l_2^m to l_2^n, where l_2^n denotes $\mathbb{R}^n$ equipped with the euclidean norm :

$$\text{For } x = (x_j)_{j \leqslant n} \in \mathbb{R}^n, \qquad \|x\|_2 = \left(\sum_{j \leqslant n} x_j^2 \right)^{\frac{1}{2}}.$$

We denote by $s_1(U) \geqslant ... \geqslant s_{n \wedge m}(U)$ the singular values of U. The operator norm of U is denoted by $\|U\|$ and is equal to $s_1(U)$. The Hilbert-Schmidt norm of U is given by

$$\|U\|_{\mathrm{HS}} = \sqrt{\mathrm{Tr}(UU^t)}.$$

We are interested in extracting columns of the matrix U in a way that the restricted matrix satisfies better properties or that the required data is reduced.

0.1.1 Restricted invertibility

An essential property of matrices is clearly their invertibility and more precisely their injectivity. In general, a matrix is not and cannot be injective in the rectangular case where the dimension of the domain is larger than that of the codomain. In this first part, we are interested in the following problem :

Given an $n \times m$ matrix U, extract a "large" number of linearly independent columns and estimate the smallest singular value of the extracted matrix.

By the linear algebra, we know that the number of linearly independent columns that can be extracted is equal to the rank of the matrix. Nevertheless, this is inconvenient, being unstable under small perturbation. To see this, take an $n \times n$ diagonal matrix D as follows

$$D = \begin{pmatrix} 1+\delta & 0 & \dots & 0 \\ 0 & \delta & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \delta \end{pmatrix}.$$

It is obvious that $D - \delta \cdot Id$ is of rank 1 while D is of rank n. Therefore, even after a small perturbation, the rank decreases from n to 1. The notion of rank does not take into consideration the fact that D lies on one eigenspace. We replace the notion of rank by what is known as the stable rank, defined by

$$\mathrm{srank}(U) = \frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2}.$$

In the previous example, one can see that if δ is small then the stable rank of D is of order 1 before and after perturbation. Note that

$$\mathrm{srank}(U) = \frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2} = \frac{\sum_{j=1}^{\mathrm{rank}(U)} s_j(U)^2}{s_1(U)^2} \leqslant \mathrm{rank}(U).$$

Thus, the "large" number of columns of U that we will seek to extract will be given by $\mathrm{srank}(U)$. At each time, $\sigma \subset \{1, ..., m\}$ will denote the set of indices of the chosen columns and U_σ the restriction of U to these columns. In other words, $U_\sigma = UP_\sigma^t$ where P_σ is the canonical projection of $\mathbb{R}^n$ into $\mathbb{R}^\sigma$. Concerning the estimation of the smallest singular value, the problem is to give a non-trivial lower bound for it. It is easy to see that

$$s_k(U) > 0 \Leftrightarrow \mathrm{rank}(U) \geqslant k,$$

thus, giving a non-trivial lower bound of the smallest singular value of U_σ means that the columns of U_σ are linearly independent which solves the problem. Also note that, giving a lower bound to the smallest singular value of U_σ is equivalent to giving an upper bound for the norm of the inverse of U_σ ; indeed,

$$\forall x \in \mathbb{R}^\sigma, \quad \|U_\sigma x\|_2^2 = \langle U_\sigma x, U_\sigma x\rangle = \left\langle U_\sigma^t U_\sigma x, x \right\rangle \geqslant s_{\min}(U_\sigma)^2 \|x\|_2^2.$$

The restricted invertibility problem was first studied by Bourgain-Tzafriri [19] who obtained a result for square matrices with norm 1 columns. Bourgain-Tzafriri proved :

Theorem 0.3 (Bourgain-Tzafriri). *Let T an $n \times n$ matrix such that $\|Te_j\|_2 = 1$, where $(e_j)_{j \leqslant n}$ denotes the canonical basis of $\mathbb{R}^n$. Then there exists $\sigma \subset \{1, .., n\}$ with*

$$|\sigma| \geqslant c \frac{n}{\|T\|^2}$$

such that for all scalars $(a_j)_{j \in \sigma}$

$$\left\| \sum_{j \in \sigma} a_j T e_j \right\|_2 \geqslant c' \left(\sum_{j \in \sigma} a_j^2 \right)^{\frac{1}{2}},$$

where c and c' are universal constants.

Since $\|Te_j\|_2 = 1$ for all $j \leqslant n$, then $\|T\|_{\mathrm{HS}}^2 = n$ and thus the size of the extracted set σ in the previous theorem is proportional to the stable rank of T. Note also that the conclusion of the theorem means that $s_{\min}(U_\sigma) \geqslant c'$. The proof given by Bourgain-Tzafriri uses random selectors, Sauer-Shelah lemma ([70], [73]) which is a combinatorial argument and finally Grothendieck's factorization theorem (see [27] et [61]). The proof is technical and non constructive. In [87], Tropp gave a randomized algorithm to achieve Grothendieck's factorization theorem and therefore, he was able to give a randomized algorithm to extract the bloc of columns given by Bourgain-Tzafriri's result. Theorem 0.3 is usually known as the Bourgain-Tzafriri restricted invertibility principle. We can also interpret the previous result as the invertibility of the operator T over the canonical decomposition of the identity $Id = \sum_{j \leqslant n} e_j e_j^t$; therefore, the problem is to search for a large part of this decomposition on which T is invertible. In this context, Vershynin [90] generalized the previous result for any decomposition of the identity and improved the size of the extraction. Let us note that Vershynin also proved a non-trivial estimate of the norm of the restricted operator (and not only of that of the inverse) ; we prefer, for now, to state a part of Vershynin's result, the one in relation with the restricted invertibility principle, while the full result will be subject of a detailed study in the next section.

Theorem 0.4 (Vershynin). *Let $Id = \sum_{j \leqslant m} x_j x_j^t$ be an identity decomposition on $\mathbb{R}^n$ and let T be a linear operator on l_2^n. For any $\varepsilon \in (0, 1)$ there exists $\sigma \subset \{1, \ldots, m\}$ of size*

$$|\sigma| \geqslant (1 - \varepsilon)\mathrm{srank}(T) = (1 - \varepsilon) \frac{\|T\|_{\mathrm{HS}}^2}{\|T\|^2}$$

such that for all scalars $(a_j)_{j\in\sigma}$

$$\left\|\sum_{j\in\sigma} a_j \frac{Tx_j}{\|Tx_j\|_2}\right\|_2 \geqslant c(\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}},$$

where $c(\varepsilon)$ *depends only on* ε*.*

Let us start comparing this result with the previous one. If, instead of an arbitrary decomposition of the identity, we take the canonical decomposition and we suppose that $\|Te_j\|_2 = 1$ then Theorem 0.4 allows us to find σ of size $(1-\varepsilon)\frac{n}{\|T\|^2}$ such that for all scalars $(a_j)_{j\in\sigma}$

$$\left\|\sum_{j\in\sigma} a_j Te_j\right\|_2 \geqslant c(\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}.$$

This recovers Theorem 0.3 with the advantage of extracting a set of indices of size almost the stable rank of T, whereas before we couldn't have more than a proportion of the stable rank. This reveals to be of a big importance since for some applications (we will see some later on), we need almost the full quantity and a proportion is not even sufficient. The constant depending on ε plays a crucial role and finding the right dependence on ε is the heart of the problem. Note also that in this result, Vershynin extracts the vectors Tx_j but the estimate deals with the normalized vectors ; this normalization is essential for some applications, especially for estimating the Banach-Mazur distance to the cube as we may see in the next part. Concerning the proof, Vershynin proves his result by a technical iteration of Theorem 0.3 combined with a result of Kashin-Tzafriri [43] which we will discuss later. Thus, the proof given by Vershynin is not constructive.

Spielman-Srivastava [74] also generalized the Bourgain-Tzafriri restricted invertibility principle. Based on the method introduced by Batson-Spielman-Srivastava [13], they proposed a deterministic algorithm to extract the bloc which performs the invertibility.

Theorem 0.5 (Spielman-Srivastava). *Let* $x_1, \dots x_m \in \mathbb{R}^n$ *such that* $Id = \sum_i x_i x_i^t$ *and let* $0 < \varepsilon < 1$*. For every linear operator* $T : \ell_2^n \to \ell_2^n$ *there exists a subset* $\sigma \subset \{1, \dots, m\}$ *of size* $|\sigma| \geqslant \left\lfloor (1-\varepsilon)^2 \frac{\|T\|_{\mathrm{HS}}^2}{\|T\|^2} \right\rfloor$ *for which* $\{Tx_i\}_{i\in\sigma}$ *is linearly independent and*

$$\lambda_{\min}\left(\sum_{i\in\sigma}(Tx_i)(Tx_i)^t\right) > \frac{\varepsilon^2\|T\|_{\mathrm{HS}}^2}{m},$$

where $\lambda_{\min}$ *is computed on* $\mathrm{span}\{Tx_i\}_{i\in\sigma}$ *or simply here* $\lambda_{\min}$ *denotes the smallest nonzero eigenvalue of the corresponding operator.*

Equivalently, for all scalars $(a_j)_{j\in\sigma}$

$$\left\|\sum_{j\in\sigma} a_j T x_j\right\|_2 \geqslant \varepsilon \frac{\|T\|_{\mathrm{HS}}}{\sqrt{m}} \left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}.$$

This result can be interpreted as restricted invertibility of rectangular matrices. Indeed, given $Id = \sum_{i\leqslant m} x_i x_i^t$ an identity decomposition on $\mathbb{R}^n$ and T a linear operator on l_2^n, we can associate to these an $n\times m$ matrix U whose columns are the vectors $(Tx_j)_{j\leqslant m}$. Since the $(x_j)_{j\leqslant m}$ form an identity decomposition, it is easy to see that $UU^t = TT^t$, and thus that $\|U\| = \|T\|$ and $\|U\|_{\mathrm{HS}} = \|T\|_{\mathrm{HS}}$. The previous result can thus be expressed in terms of the rectangular matrix U.

As a conclusion, we are facing two results generalizing the Bourgain-Tzafriri restricted invertibility principle ; on the one hand, Vershynin's result where the chosen vectors are normalized but the dependence on ε is not satisfactory, and on the other hand Spielman-Srivastava's result which provides a good dependence on ε but where the vectors are not normalized. A natural question is to find a result which includes the two previous ones in order to get the normalizations with a good dependence on ε.

We prove a restricted invertibility principle for any rectangular matrix and for any choice of normalization while keeping a good dependence on ε.

Theorem 1. *Given an* $n\times m$ *matrix* U *and a diagonal* $m\times m$ *matrix* D *with* $(\alpha_j)_{j\leqslant m}$ *on its diagonal, with the property that* $\mathrm{Ker}(D)\subset\mathrm{Ker}(U)$*, then for any* $\varepsilon\in(0,1)$ *there exists* $\sigma\subset\{1,...,m\}$ *with*

$$|\sigma| \geqslant \left\lfloor (1-\varepsilon)^2 \mathrm{srank}(U)\right\rfloor = \left\lfloor (1-\varepsilon)^2 \frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2}\right\rfloor$$

such that

$$s_{\min}\left(U_\sigma D_\sigma^{-1}\right) > \frac{\varepsilon\|U\|_{\mathrm{HS}}}{\|D\|_{\mathrm{HS}}},$$

where $s_{\min}$ *denotes the smallest singular value.*

Equivalently, for all scalars $(a_j)_{j\in\sigma}$

$$\left\|\sum_{j\in\sigma} a_j \frac{Ue_j}{\alpha_j}\right\|_2 \geqslant \varepsilon \frac{\|U\|_{\mathrm{HS}}}{\|D\|_{\mathrm{HS}}} \left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}.$$

To be convinced that Theorem 1 is not a consequence of Theorem 0.5, just note that the important point in the previous statement is that the size of the restriction depends only on the matrix U and not on the normalizing matrix D. Theorem 1 bridges the gap between the previous results and generalizes both Theorem 0.4 and Theorem 0.5. Let $Id = \sum_{j\leqslant m} x_j x_j^t$ be an identity decomposition on $\mathbb{R}^n$ and T be a linear operator on l_2^n. Define U the $n\times m$ matrix whose columns are the vectors $(Tx_j)_{j\leqslant m}$. Thus, as already noted before, we have $UU^t = TT^t$, and applying Theorem 1 we find $\sigma \subset \{1, ..., m\}$ of size

$$|\sigma| \geqslant (1-\varepsilon)^2 \mathrm{srank}(U) = (1-\varepsilon)^2 \mathrm{srank}(T).$$

Taking $D = Id$, we get for any scalars $(a_j)_{j\in\sigma}$

$$\left\| \sum_{j\in\sigma} a_j T x_j \right\|_2 = \left\| \sum_{j\in\sigma} a_j U e_j \right\|_2 \geqslant \varepsilon \frac{\|U\|_{\mathrm{HS}}}{\|D\|_{\mathrm{HS}}} \left(\sum_{j\in\sigma} a_j^2 \right)^{\frac{1}{2}} = \varepsilon \frac{\|T\|_{\mathrm{HS}}}{\sqrt{m}} \left(\sum_{j\in\sigma} a_j^2 \right)^{\frac{1}{2}},$$

and we recover Theorem 0.5. Taking $D = diag(\alpha_1, ..., \alpha_m)$ with $\alpha_j = \|Tx_j\|_2$, we get for any scalars $(a_j)_{j\in\sigma}$

$$\left\| \sum_{j\in\sigma} a_j \frac{Tx_j}{\|Tx_j\|_2} \right\|_2 = \left\| \sum_{j\in\sigma} a_j \frac{Ue_j}{\alpha_j} \right\|_2 \geqslant \varepsilon \frac{\|U\|_{\mathrm{HS}}}{\|D\|_{\mathrm{HS}}} \left(\sum_{j\in\sigma} a_j^2 \right)^{\frac{1}{2}} = \varepsilon \left(\sum_{j\in\sigma} a_j^2 \right)^{\frac{1}{2}},$$

and we recover Theorem 0.4.

0.1.2 Norm of coordinate restrictions

In the preceding, we studied the invertibility problem which concerns only the smallest singular value. In this section, we will only deal with the largest singular value which represents the norm of the matrix. Given an $n\times m$ matrix U and an integer $k \leqslant m$, we are interested in extracting k columns of U which minimize the operator norm. This problem was first studied in [51] and then by Kashin-Tzafriri [43]. Kashin-Tzafriri ([43], see also [90]) proved :

Theorem 0.6 (Kashin-Tzafriri). *Let U be an $n\times m$ matrix. Fix λ with $1/m \leqslant \lambda \leqslant \frac{1}{4}$. Then, there exists a subset $\nu \subset \{1, \dots, m\}$ of cardinality $|\nu| \geqslant \lambda m$ such that*

$$\|U_\nu\| \leqslant c \left(\sqrt{\lambda} \|U\|_2 + \frac{\|U\|_{\mathrm{HS}}}{\sqrt{m}} \right),$$

where $U_\nu = UP_\nu^t$ and P_ν denotes the coordinate projection onto $\mathbb{R}^\nu$.

The conclusion of the theorem means that for $\lambda \leqslant \frac{1}{4}$ fixed we have

$$\min_{\substack{\sigma\subset\{1,\ldots,m\}\\ |\sigma|=\lambda m}} \|U_\sigma\| \leqslant c\left(\sqrt{\lambda}\|U\| + \frac{\|U\|_{\mathrm{HS}}}{\sqrt{m}}\right) \tag{3}$$

This result is optimal (up to a constant) in the sense that the right hand side of the estimate cannot be replaced with a smaller quantity. To be convinced, consider the following example :

Let $m, h, k \in \mathbb{N}$ such that $m = hk$. We divide the set $\{1, ..., m\}$ into h disjoint sets I_l of size k each. Let U be an $m \times m$ matrix defined by $Ue_j = \frac{1}{\sqrt{k}}e_l$ if $j \in I_l$ for $l \leqslant h$, where $(e_j)_{j\leqslant m}$ denotes the canonical basis of $\mathbb{R}^m$. Thus, we have

$$UU^t = \sum_{j\leqslant m} (Ue_j)(Ue_j)^t = \sum_{l\leqslant h}\sum_{j\in I_l} \frac{1}{k} e_l e_l^t = \sum_{l\leqslant h} e_l e_l^t$$

Therefore, $\|U\| = 1$ and $\|U\|_{\mathrm{HS}} = \sqrt{h}$. Now let p be the number of columns we want to extract (p plays the role of λm in Theorem 0.6). If $p \leqslant h$, then the best restriction would be to choose only one vector per block and in that case, the norm of the restriction is $\frac{1}{\sqrt{k}} = \sqrt{\frac{h}{m}}$ which is the second term in (3). If $p > h$ then the best restriction would be to choose as much vectors in different blocks as possible, which means choosing $\frac{p}{h}$ vectors from each block. In that case, the norm of the restriction is $\sqrt{\frac{p}{hk}} = \sqrt{\frac{p}{m}}$ which plays the role of the first term in (3).

The proof of Theorem 0.6 ([43], see also [90]) uses random selectors and Grothendieck's factorization theorem (see [27] and [61]). The method is not constructive. In [87], Tropp gave a randomized algorithm to achieve Grothendieck's factorization theorem and therefore he was able to give a randomized algorithm to find the set σ given by Theorem 0.6. Our aim is to give a new proof of this result using tools from the method of Batson-Spielman-Srivastava [13].

We obtain a result which improves the size of the extracted matrix and all the constants appearing in Theorem 0.6. An important point is that our proof provides a deterministic algorithm to achieve the extraction.

Theorem 2. *Let U be an $n \times m$ matrix and let $1/m \leqslant \lambda \leqslant \eta < 1$. Then, there exists $\sigma \subset \{1, \ldots, m\}$ with $|\sigma| = k \geqslant \lambda m$ such that*

$$\|U_\sigma\| \leqslant \frac{1}{\sqrt{1-\lambda}}\left(\sqrt{\lambda+\eta}\|U\| + \sqrt{1+\frac{\lambda}{\eta}}\frac{\|U\|_{\mathrm{HS}}}{\sqrt{m}}\right),$$

In particular,

$$\|U_\sigma\| \leqslant \frac{\sqrt{2}}{\sqrt{1-\lambda}} \left(\sqrt{\lambda}\|U\| + \frac{\|U\|_{\mathrm{HS}}}{\sqrt{m}} \right)$$

0.1.3 Extracting a square block : first attempt

Until now, we always managed to extract a block of columns of a fixed matrix. Given an $n \times n$ square matrix T with zero diagonal, Bourgain-Tzafriri ([19],[20]) worked on extracting a "large" square submatrix with small norm. They proved the following :

Theorem 0.7. *There is a universal constant $c > 0$ such that for every $\varepsilon > 0$ and $n \in \mathbb{N}$, if an operator $T : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $\langle Te_i, e_i \rangle = 0$ for all $i \in \{1, \dots, n\}$ then there exists a subset $\sigma \subseteq \{1, \dots, n\}$ satisfying $|\sigma| \geqslant c\varepsilon^2 n$ and $\|P_\sigma T P_\sigma^t\| \leqslant \varepsilon \|T\|$.*

It should be noted that Theorem 0.7 implies the Bourgain-Tzafriri restricted invertibility principle (Theorem 0.3) but with a bad dependence on the norm of T. To see this, take T an $n \times n$ matrix such that $\|Te_j\|_2 = 1$ for any $j \leqslant n$. Denoting $A = T^tT - Id$, then A has zero diagonal. Apply Theorem 0.7 to find $\sigma \subset \{1, ..., n\}$ of size $c\varepsilon^2 n$ such that $\|P_\sigma A P_\sigma^t\| \leqslant \varepsilon \|A\|$. This means that

$$-\varepsilon \|A\| \cdot Id \preceq P_\sigma A P_\sigma^t \preceq \varepsilon \|A\| \cdot Id$$

Noting that $\|T\| \geqslant 1$ and thus that $\|A\| \leqslant 2\|T\|^2$, we get

$$(1 - 2\varepsilon \|T\|^2) \cdot Id \preceq P_\sigma T^t T P_\sigma^t \preceq (1 + 2\varepsilon \|T\|^2) \cdot Id$$

Taking $\varepsilon = \frac{1}{4\|T\|^2}$, we found σ of size $c\frac{n}{\|T\|^4}$ such that $s_{\min}(T_\sigma) \geqslant \frac{1}{2}$. This recovers Theorem 0.3 with a bad dependence on the norm of T.

In [14], it is proven that the quadratic dependence on ε in Theorem 0.7 is optimal. Indeed, take A an $n \times n$ Hadamard matrix i.e a matrix whose columns contain only ± 1 and are mutually orthogonal. We have $AA^t = nId$, and thus $\|A\| = \sqrt{n}$. If P is a canonical projection of rank k i.e P is a diagonal matrix having k terms equal to 1 on its diagonal, then PAP^t is a $k \times k$ matrix containing only ± 1. Therefore $\|PAP^t\| \geqslant \sqrt{k}$ and we have

$$\|P\left(A - \mathrm{diag}(A)\right)P^t\| \geqslant \|PAP^t\| - \|P\mathrm{diag}(A)P^t\| \geqslant \sqrt{k} - 1 \geqslant \frac{\sqrt{k}}{2},$$

if $k \geqslant 4$ for example. As a conclusion, if we note $B = \frac{1}{\sqrt{n}}\left(A - \mathrm{diag}(A)\right)$ where A is an $n \times n$

Hadamard matrix then for any set of indices σ of size $k = \varepsilon^2 n$, we have

$$\|P_\sigma B P_\sigma^t\| \geqslant \frac{\sqrt{k}}{2\sqrt{n}} = \frac{\varepsilon}{2} \geqslant \frac{\varepsilon}{4}\|B\|,$$

since $\|B\| \leqslant 2$. This shows that the quadratic dependence on ε in Theorem 0.7 is needed.

In a survey [55] on the method of Batson-Spielman-Srivastava [13], Naor asked about giving a new proof of Theorem 0.7 using the method of Batson-Spielman-Srivastava. We were then interested in this problem and to begin, we present a first attempt which will not give the right dependence on ε.

If T is an $n \times n$ symmetric matrix, write $U = \left[\frac{1}{\|T\|}\left(T + \|T\| \cdot Id\right)\right]^{\frac{1}{2}}$. In order to prove that $\|P_\sigma T P_\sigma^t\| \leqslant \varepsilon\|T\|$, it is sufficient to prove that

$$(1-\varepsilon) \cdot Id \preceq P_\sigma U^t U P_\sigma^t \preceq (1+\varepsilon) \cdot Id$$

The main idea is to run the algorithm of Theorem 2 to find a set of indices ν satisfying the upper bound, then run the algorithm of Theorem 1 in order to find a set of indices σ inside ν satisfying the lower bound. This procedure allows us to prove the following :

Proposition 1. *Let T be an $n \times n$ symmetric matrix with zero diagonal. There exists $\sigma \subset \{1, .., n\}$ of cardinality $c\varepsilon^4 n$ such that $\|P_\sigma T P_\sigma^t\| \leqslant \varepsilon\|T\|$.*

Therefore, our first attempt produces a deterministic algorithm to prove Theorem 0.7 but with a bad dependence on ε and with the assumption of symmetry on the matrix T. Running each algorithm separately produces this error in the dependence on ε. We will be interested in finding a method which allows us to obtain a lower and an upper bound for the singular values of the restricted matrix.

0.1.4 Extracting a well conditioned block

In the first part, we mentioned a part of a result due to Vershynin [90] which we stated in Theorem 0.4. The full statement is the following :

Theorem 0.8 (Vershynin). *Let $Id = \sum_{j \leqslant m} x_j x_j^t$ and let T be a linear operator on ℓ_2^n. For any $\varepsilon \in (0,1)$, one can find $\sigma \subset \{1, \ldots, m\}$ with*

$$|\sigma| \geqslant \lfloor (1-\varepsilon)\text{srank}(T) \rfloor = \left\lfloor (1-\varepsilon)\frac{\|T\|_{\text{HS}}^2}{\|T\|^2} \right\rfloor$$

such that for all scalars $(a_j)_{j\in\sigma}$

$$c_1(\varepsilon)\left(\sum_{j\in\sigma}a_j^2\right)^{\frac{1}{2}} \leqslant \left\|\sum_{j\in\sigma}a_j\frac{Tx_j}{\|Tx_j\|_2}\right\|_2 \leqslant c_2(\varepsilon)\left(\sum_{j\in\sigma}a_j^2\right)^{\frac{1}{2}},$$

where $c_1(\varepsilon), c_2(\varepsilon)$ *depend only on* ε.

The dependence on ε satisfies $c(\varepsilon) = \frac{c_2(\varepsilon)}{c_1(\varepsilon)} \approx \varepsilon^{c\log(\varepsilon)}$. In other terms, the previous result states that one can find a "large" part (say of size k) of the sequence $(Tx_j)_{j\leqslant m}$ which is $c(\varepsilon)$-equivalent to an orthogonal basis of l_2^k. Vershynin applies this to the study of contact points and embeddings of the cube in an arbitrary Banach space ; indeed, his result combined with results of Talagrand [83] gives a clear picture on the form of these embeddings. In Theorem 1, we improved a part of Theorem 0.8 ; precisely, the lower bound dealing with the restricted invertibility principle. Obviously, our target is to improve the full result of Vershynin. Let us start formulating the result in terms of matrices :

Given $Id = \sum_{j\leqslant m} x_j x_j^t$ an identity decomposition on $\mathbb{R}^n$ and T a linear operator on l_2^n, consider the $n\times m$ matrix U having the vectors $(Tx_j)_{j\leqslant m}$ as columns. We denote by $\tilde{U}$ the matrix U with normalized columns. It is clear that $UU^t = TT^t$, and thus Theorem 0.8 implies the existence of a set of indices σ of size $(1-\varepsilon)\mathrm{srank}(U)$ such that

$$c_1(\varepsilon) \leqslant s_{\min}(\tilde{U}_\sigma) \leqslant s_{\max}(\tilde{U}_\sigma) \leqslant c_2(\varepsilon).$$

The conditioning number of a matrix U is given by

$$\kappa(U) = \max\left\{\frac{\|Ux\|_2}{\|Uy\|_2};\ \|x\|_2 = \|y\|_2 = 1\right\} = \frac{s_{\max}(U)}{s_{\min}(U)}$$

Obviously, if the matrix is not of full rank then its conditioning number is infinite. An interesting problem is to extract a big block of the matrix such that the conditioning number of the restriction is bounded. If the conditioning number is close to 1, then the matrix is a multiple of an isometry. Theorem 0.8 allows to extract from $\tilde{U}$ a number of columns equal to $(1-\varepsilon)\mathrm{srank}(U)$ such that $\kappa(\tilde{U}_\sigma) \leqslant \varepsilon^{c\log(\varepsilon)}$.

In order to improve Vershynin's result, the idea is to merge the algorithms of Theorem 1 and Theorem 2 to get simultaneously a restricted invertibility principle (and thus an estimate of $s_{\min}$) and an estimate of $s_{\max}$. We proved the following :

Theorem 3. *Let U be an $n \times m$ matrix and denote by $\widetilde{U}$ the matrix whose columns are the columns of U normalized. For any $\varepsilon \in (0,1)$, there exists $\sigma \subset \{1,...,m\}$ of size*

$$|\sigma| \geqslant \left\lfloor (1-\varepsilon)^2 \text{srank}(U) \right\rfloor = \left\lfloor (1-\varepsilon)^2 \frac{\|U\|_{\text{HS}}^2}{\|U\|^2} \right\rfloor$$

such that

$$\frac{\varepsilon}{2-\varepsilon} \leqslant s_{\min}\left(\widetilde{U}_\sigma\right) \leqslant s_{\max}\left(\widetilde{U}_\sigma\right) \leqslant \frac{2-\varepsilon}{\varepsilon}$$

In other terms, for all scalars $(a_j)_{j\in\sigma}$

$$\frac{\varepsilon}{2-\varepsilon}\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}} \leqslant \left\|\sum_{j\in\sigma} a_j \frac{Ue_j}{\|Ue_j\|_2}\right\|_2 \leqslant \frac{2-\varepsilon}{\varepsilon}\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}.$$

Therefore, our result implies the existence of a block of size $(1-\varepsilon)^2\text{srank}(U)$ whose conditioning number is less than $\left(\frac{2-\varepsilon}{\varepsilon}\right)^2$. In the regime where ε is close to 1, Theorem 3 allows to extract an almost isometric block.

Corollary 1. *Let U be an $n \times m$ matrix and denote by $\widetilde{U}$ the matrix whose columns are the columns of U normalized. For any $\varepsilon \in (0,1)$, there exists $\sigma \subset \{1,...,m\}$ of size*

$$|\sigma| \geqslant \left\lfloor \frac{\varepsilon^2}{9} . \frac{\|U\|_{\text{HS}}^2}{\|U\|^2} \right\rfloor$$

such that

$$1-\varepsilon \leqslant s_{\min}\left(\widetilde{U}_\sigma\right) \leqslant s_{\max}\left(\widetilde{U}_\sigma\right) \leqslant 1+\varepsilon$$

In other terms, for all scalars $(a_j)_{j\in\sigma}$

$$(1-\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}} \leqslant \left\|\sum_{j\in\sigma} a_j \frac{Ue_j}{\|Ue_j\|_2}\right\|_2 \leqslant (1+\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}.$$

Later, we will see that Corollary 1 implies (in the case of a symmetric operator) Theorem 0.7 which answers Naor's question. As we have already seen, the dependence on ε is optimal in Theorem 0.7 which means that, up to a constant, Corollary 1 is optimal.

0.1.5 Column paving and the Kadison-Singer conjecture

An $n \times m$ matrix U is called standardized if all its columns are of norm 1. Note that the Bourgain-Tzafriri restricted invertibility principle (Theorem 0.3) dealt with standardized matrices. The column paving problem consists of partitioning a matrix into well conditioned blocks. Precisely, we would like to give an algorithm to partition a matrix into almost isometric blocks. A first result in this direction is due to Bourgain-Tzafriri ([19],[20]) but is not constructive. Tropp gave a randomized algorithm to achieve the partition into almost isometric blocks (see [86] for further information).

Using Corollary 1, we give a deterministic algorithm to recover a result of Bourgain-Tzafriri and achieve the column paving while improving all the constants involved.

Proposition 2. *Let U be an $n \times m$ standardized matrix. For any $\varepsilon \in (0,1)$, there exists a partition of $\{1, ..., m\}$ into p sets $\sigma_1, ..., \sigma_p$ such that*

$$p \leqslant \frac{9\|U\|^2 \log(m)}{\varepsilon^2}$$

and for any $i \leqslant p$,

$$1 - \varepsilon \leqslant s_{\min}\left(U_{\sigma_i}\right) \leqslant s_{\max}\left(U_{\sigma_i}\right) \leqslant 1 + \varepsilon$$

Let us now return to the problem of extracting a square submatrix. In our first attempt, we recovered Theorem 0.7 but with a bad dependence on ε. Having launched the algorithms of Theorem 2 and Theorem 1 consecutively was the cause of the error in the estimate. Now that we have an algorithm merging the two previous ones, we recover Theorem 0.7 with the right dependence on ε in the case of symmetric matrices. Indeed, if T is an $n \times n$ symmetric matrix, we note $U = \left[\frac{1}{\|T\|}\left(T + \|T\| \cdot Id\right)\right]^{\frac{1}{2}}$ and thus to prove that $\|P_\sigma T P_\sigma^t\| \leqslant \varepsilon\|T\|$, it is sufficient to show that

$$(1 - \varepsilon) \cdot Id \preceq P_\sigma U^t U P_\sigma^t \preceq (1 + \varepsilon) \cdot Id,$$

which is already established in Corollary 1.

Following the procedure described above, we obtain the following :

Proposition 3. *Let T be an $n \times n$ symmetric matrix with 0 diagonal. For any $\varepsilon \in (0,1)$, there exists $\sigma \subset \{1, ..., n\}$ of size*

$$|\sigma| \geqslant \frac{(\sqrt{2}-1)^4 \varepsilon^2 n}{2}$$

such that

$$\left\|P_\sigma T P_\sigma^t\right\| \leqslant \varepsilon\|T\|$$

The previous result is important because it is directly related to the Kadison-Singer conjecture [41]. This conjecture is still unsolved and many equivalent formulations were proposed ([7], see [22] and [23] for more details). We present here what is known as the "Paving conjecture" due to Anderson [7] and which is equivalent to Kadison-Singer.

Conjecture. *For any $\varepsilon \in (0,1)$, there exists $p = p(\varepsilon)$ such that for any $n \in \mathbb{N}$ and any $n \times n$ matrix T with zero diagonal, there exists a partition of $\{1, ..., n\}$ into p sets $\sigma_1, ..., \sigma_p$ such that*

$$\forall i \leqslant p, \quad \|P_{\sigma_i} T P_{\sigma_i}^t\| \leqslant \varepsilon \|T\|$$

Note that it is sufficient to prove the conjecture for symmetric matrices (see [22]). At present, the two strongest results on the paving problem are due to Bourgain-Tzafriri [20]. The first shows that there exists a partition of size of order $\log(n)$ for which the conclusion of the conjecture holds. The second states that the conjecture is true for matrices whose entries are uniformly bounded by a suitable quantity.

By iterating Proposition 3, we give a deterministic algorithm, which improves all the constants involved in the result, to partition a symmetric matrix with zero diagonal, into a number of blocks of order $\log(n)$ for which the conclusion of the conjecture holds true.

Proposition 4. *Let T be an $n \times n$ symmetric matrix with 0 diagonal. For any $\varepsilon \in (0,1)$, there exists a partition of $\{1, ..., n\}$ into k subsets $\sigma_1, .., \sigma_k$ such that*

$$k \leqslant \frac{2\log(n)}{(\sqrt{2}-1)^4 \varepsilon^2}$$

and for any $i \leqslant k$,

$$\left\| P_{\sigma_i} T P_{\sigma_i}^t \right\| \leqslant \varepsilon \|T\|$$

0.1.6 Application to harmonic analysis

Denote by $\mathbb{T}$ the circle and ν the normalized Lebesgue measure on $\mathbb{T}$. The set of functions of square integrable on $\mathbb{T}$ is denoted by $L_2(\mathbb{T}, \nu)$. For any $f \in L_2(\mathbb{T}, \nu)$, define

$$\|f\|_{L_2(\mathbb{T})} = \left(\int_{\mathbb{T}} |f|^2 d\nu \right)^{\frac{1}{2}}.$$

For any subset B of $\mathbb{T}$, define

$$\|f\|_{L_2(B)} = \left(\frac{1}{\nu(B)}\int_{\mathbb{B}} |f|^2 d\nu\right)^{\frac{1}{2}}.$$

If Λ is a set of integers, the density of Λ is given by :

$$\text{dens}(\Lambda) = \lim_{n\to\infty} \frac{|\Lambda \cap \{-n, n\}|}{2n},$$

whenever the limit exists.
We also note $L_2^\Lambda(\mathbb{T},\nu)$ the subspace of $L_2(\mathbb{T},\nu)$ spanned by $\{e^{i.kx}\}_{k\in\Lambda}$. In other words, $L_2^\Lambda(\mathbb{T},\nu)$ represents the set of functions whose Fourier transforms are supported on Λ.

In [19], Bourgain-Tzafriri gave an application of the restricted invertibility principle to harmonic analysis by proving :

Theorem 0.9 (Bourgain-Tzafriri)**.** *For any $B \subset \mathbb{T}$, one can find a subset Λ of the integers with* $\text{dens}(\Lambda) \geqslant c\nu(B)$*, such that for any $f \in L_2^\Lambda(\mathbb{T},\nu)$ we have*

$$\|f\|_{L_2(B)} \geqslant c' \cdot \|f\|_{L_2(\mathbb{T})}, \tag{4}$$

where c, c' are universal constants.

In [89], Vershynin uses Theorem 0.8 to give also an upper bound in (4). Using Theorem 3, we improve the results obtained by Bourgain-Tzafriri [19] and Vershynin [89].

Theorem 4. *Let B be a subset of $\mathbb{T}$ of positive measure. For any $\varepsilon \in (0,1)$, there exists a set of integers Λ with density* $\text{dens}(\Lambda) \geqslant (1-\varepsilon)^2\nu(B)$ *such that for any $f \in L_2^\Lambda(\mathbb{T},\nu)$, we have*

$$\frac{\varepsilon}{2-\varepsilon}\|f\|_{L_2(\mathbb{T})} \leqslant \|f\|_{L_2(B)} \leqslant \frac{2-\varepsilon}{\varepsilon}\|f\|_{L_2(\mathbb{T})}$$

In the regime where ε is close to 1, we get the following corollary :

Corollary 2. *Let B be a subset of $\mathbb{T}$ of positive measure. For any $\varepsilon \in (0,1)$, there exists a set of integers Λ with density* $\text{dens}(\Lambda) \geqslant \frac{\varepsilon^2}{9}\nu(B)$ *such that for any $f \in L_2^\Lambda(\mathbb{T},\nu)$, we have*

$$(1-\varepsilon)\|f\|_{L_2(\mathbb{T})} \leqslant \|f\|_{L_2(B)} \leqslant (1+\varepsilon)\|f\|_{L_2(\mathbb{T})}$$

This means that for any $B \subset \mathbb{T}$ with positive measure, one can find a set of integers Λ with positive density such that the two norms $\|\cdot\|_{L_2(\mathbb{T})}$ and $\|\cdot\|_{L_2(B)}$ are equivalent on $L_2^\Lambda(\mathbb{T},\nu)$ with an equivalence factor close to 1.

0.2 Banach-Mazur distance to the cube

Studying distance between objects is an effective tool towards their identification. In each object category, one must define a notion of distance which is coherent with that of applications defined on this category. In the setting of Banach spaces, we are interested in the Banach-Maazur distance which will measure how two spaces are isomorphic. To have a clear idea, consider an n-dimensional Banach space as $\mathbb{R}^n$ equipped with a norm ; measuring the geometric distance between two Banach spaces is equivalent to measuring the degree of equivalence between their norms. It is known that two norms $|\cdot|$ and $\|\cdot\|$ on $\mathbb{R}^n$ are equivalent i.e

$$\alpha|\cdot| \leqslant \|\cdot\| \leqslant \beta|\cdot|,$$

but the question is to estimate the equivalence factor $\frac{\beta}{\alpha}$ which can depend on the dimension. Thus, if we estimate the equivalence factor between the norms of two Banach spaces X and Y, we get

$$\alpha B_X \subset B_Y \subset \beta B_X,$$

where B_X and B_Y denote the unit balls of X and Y respectively ; the geometric distance between X and Y is the equivalence factor between their norms. In order to calculate the Banach-Mazur distance between X and Y, one should take the minimum, over all $T \in GL_n(\mathbb{R})$, of the geometric distance between X and TY. We will denote by $\mathcal{BM}_n$ the Banach-Mazur compactum i.e the set of n-dimensional Banach spaces. The Banach-Mazur distance between two elements X and Y of $\mathcal{BM}_n$ is given by :

$$d(X,Y) = \inf\{\|T\| \cdot \|T^{-1}\| \mid \ T \text{ is an isomorphism between } X \text{ and } Y\ \}$$

If X and Y are isometric, then the distance is equal to 1. It is easy to check that d is multiplicative, which means that

$$\forall X, Y, Z \in \mathcal{BM}_n, \quad d(X,Z) \leqslant d(X,Y)d(Y,Z).$$

Thus $\log(d)$ is a distance over the quotient of $\mathcal{BM}_n$ obtained by identifying isometric spaces. Another important property is that d is invariant by duality i.e

$$\forall X, Y \in \mathcal{BM}_n, \quad d(X,Y) = d(X^*, Y^*).$$

Estimate the distance between two Banach spaces is very important in order to understand the structure of $\mathcal{BM}_n$. The first step is obviously to look at examples and more precisely, estimate the distance between the spaces l_p^n whose norms are given by

$$\|x\|_p = \left(\sum_{j\leqslant n} |x_j|^p\right)^{\frac{1}{p}} \text{ if } 1 \leqslant p < \infty \quad \text{and} \quad \|x\|_\infty = \max_{j\leqslant n} |x_j|,$$

where $x = (x_j)_{j\leqslant n} \in \mathbb{R}^n$. It is easy to check that $d(l_1^n, l_2^n) = d(l_\infty^n, l_2^n) = \sqrt{n}$ and more generally that $d(l_p^n, l_q^n) = n^{\frac{1}{p}-\frac{1}{q}}$ if $1 \leqslant p \leqslant q \leqslant 2$ or $2 \leqslant p \leqslant q \leqslant \infty$. When $1 \leqslant p < 2 < q \leqslant \infty$, it is proven in [37] that

$$\frac{1}{\sqrt{2}} \max\left(n^{\frac{1}{p}-\frac{1}{2}}, n^{\frac{1}{2}-\frac{1}{q}}\right) \leqslant d(l_p^n, l_q^n) \leqslant (1+\sqrt{2}) \max\left(n^{\frac{1}{p}-\frac{1}{2}}, n^{\frac{1}{2}-\frac{1}{q}}\right),$$

and in particular $d(l_1^n, l_\infty^n)$ is of order $\sqrt{n}$. Apart from few examples and special cases, estimate the Banach-Mazur distance in full generality appears to be a difficult problem. John's result [40] which we stated in Theorem 0.2 allows to prove that the distance between any element of $\mathcal{BM}_n$ and l_2^n is at most $\sqrt{n}$. If we note

$$R_2^n = \max\left\{d(X, l_2^n)/\ X \in \mathcal{BM}_n\right\},$$

then John's theorem states that $R_2^n \leqslant \sqrt{n}$ and since $d(l_1^n, l_2^n) = \sqrt{n}$ then $R_2^n = \sqrt{n}$. By multiplicativity of the distance, we have that $d(X,Y) \leqslant n$ for any $X, Y \in \mathcal{BM}_n$ and thus that the diameter of $\mathcal{BM}_n$ is less or equal to n. Gluskin [34] showed that n is the right order for the diameter of $\mathcal{BM}_n$ by proving the existence of two spaces X and Y such that $d(X,Y) \geqslant cn$. Since the diameter of $\mathcal{BM}_n$ is of order n and $R_2^n = \sqrt{n}$ then l_2^n is a center of the Banach-Mazur compactum $\mathcal{BM}_n$.

0.2.1 Upper bound for the Banach-Mazur distance to the cube

In this thesis, we were interested in the estimate of the Banach-Mazur distance to l_∞^n whose unit ball is the n-dimensional cube. We have already seen that the distance from the spaces l_p^n to l_∞^n, and more surprisingly from l_1^n to l_∞^n, does not exceed $\sqrt{n}$ asymptotically. A natural question is to ask whether l_∞^n is also a center of $\mathcal{BM}_n$ and what is the distance from an arbitrary Banch space to l_∞^n. By a similar construction to Gluskin's one, Szarek [80] proved the existence of a Banach space X such that $d(X, l_\infty^n) \geqslant c\sqrt{n}\log(n)$, which means that l_∞^n is not a center of $\mathcal{BM}_n$. Note

$$R_\infty^n = \max\left\{d(X, l_\infty^n)/\ X \in \mathcal{BM}_n\right\} \quad \text{and} \quad R_1^n = \max\left\{d(X, l_1^n)/\ X \in \mathcal{BM}_n\right\}$$

By duality, one can write $R_\infty^n = R_1^n$ and by the result of Szarek [80], we have $R_1^n \geqslant c\sqrt{n}\log(n)$. The problem of giving an upper bound for R_∞^n was first studied by Bourgain-Szarek [18] who proved that $R_\infty^n = o(n)$. Szarek-Talagrand [82] then Giannopoulos [31] improved this estimate to obtain $cn^{\frac{7}{8}}$ and $cn^{\frac{5}{6}}$ respectively. These proofs are based on Dvoretzky-Rogers type factorization which will be discussed in the next section. Note also that Taschuk [84] established an estimate of R_∞^n for small dimensions ; precisely,

$$R_\infty^n \leqslant \sqrt{n^2 - 2n + 2 + \frac{2}{\sqrt{n+2} - 1}},$$

which is not satisfactory for big dimensions since it is of order n.

We propose a new proof of what is called "proportional Dvoretzky-Rogers factorization" and which is the heart of the previous methods. This will be detailed later, but this result allows us to improve the constants involved in the final estimate of R_∞^n while recovering the right asymptotic behavior.

Theorem 5. *Let X be an n-dimensional Banach space. Then*

$$d(X, l_1^n) \leqslant 2^{\frac{4}{3}}\sqrt{n} \cdot d(X, l_2^n)^{\frac{2}{3}}.$$

Moreover, $R_\infty^n \leqslant (2n)^{\frac{5}{6}}$.

Note also that this estimate improves Taschuk's one for dimensions bigger than 22.

0.2.2 "Proportional Dvoretzky-Rogers factorization" : symmetric case

The Dvoretzky-Rogers lemma [28] states that if X is an n-dimensional Banach-space then there exist $x_1, ..., x_k \in X$ with $k = \sqrt{n}$ such that for all scalars $(a_j)_{j\leqslant k}$, we have

$$\max_{j\leqslant k} |a_j| \leqslant \left\| \sum_{j\leqslant k} a_j x_j \right\|_X \leqslant c \left(\sum_{j\leqslant k} a_j^2 \right)^{\frac{1}{2}},$$

where c is a universal constant. Bourgain-Szarek [18] showed that this result holds for k proportional to n, this is why it is called "the proportional Dvoretzky-Rogers factorization".

Theorem 0.10 (Proportional Dvoretzky-Rogers factorization)**.** *Let X be an n-dimensional Banach space. For any $\varepsilon \in (0,1)$, there exist $x_1, ..., x_k \in X$ with $k \geqslant (1-\varepsilon)n$ such that for all* *scalars* $(a_j)_{j\leqslant k}$

$$\max_{j\leqslant k}|a_j| \leqslant \left\|\sum_{j\leqslant k} a_j x_j\right\|_X \leqslant c(\varepsilon)\left(\sum_{j\leqslant k} a_j^2\right)^{\frac{1}{2}},$$

where $c(\varepsilon)$ *is a constant depending on* ε. *Equivalently, the identity operator* $i_{2,\infty} : l_2^k \longrightarrow l_\infty^k$ *can be written* $i_{2,\infty} = \alpha \circ \beta$ *with* $\beta : l_2^k \longrightarrow X, \alpha : X \longrightarrow l_\infty^k$ *and* $\|\alpha\| \cdot \|\beta\| \leqslant c(\varepsilon)$.

Finding the right dependence on ε is an important problem and remains open till now. Szarek [80] showed that one cannot hope for a dependence better than $c\varepsilon^{-\frac{1}{10}}$. Szarek-Talagrand [82] proved Theorem 0.10 with $c(\varepsilon) = c\varepsilon^{-2}$ while Giannopoulos improved the estimate to get $c\varepsilon^{-\frac{3}{2}}$ and $c\varepsilon^{-1}$ in [31] and [32] respectively. In all these results, a factorization of the identity $i_{1,2}$ was established, the factorization of $i_{2,\infty}$ following by duality. The previous proofs used geometric arguments, including Dvoretzky-Rogers lemma, and some technical combinatorics alongside Grothendieck's factorization theorem.

We propose a basic approach based on the normalized restricted invertiblety principle which we proved in Theorem 1. This substantially simplify the proof and allows us to get the following :

Theorem 6. *Let* X *be an* n*-dimensional Banach space. For any* $\varepsilon \in (0,1)$, *there exist* $x_1, ..., x_k \in X$ *with* $k \geqslant [(1-\varepsilon)^2 n]$ *such that for all scalars* $(a_j)_{j\leqslant k}$

$$\varepsilon\left(\sum_{j\leqslant k} a_j^2\right)^{\frac{1}{2}} \leqslant \left\|\sum_{j\leqslant k} a_j x_j\right\|_X \leqslant \sum_{j\leqslant k}|a_j|$$

Equivalently, the identity operator $i_{1,2} : l_1^k \longrightarrow l_2^k$ *can be written as* $i_{1,2} = \alpha \circ \beta$, *where* $\beta : l_1^k \longrightarrow X$, $\alpha : X \longrightarrow l_2^k$ *and* $\|\alpha\| \cdot \|\beta\| \leqslant \varepsilon^{-1}$.

As a direct application of the previous result, we have the following corollary :

Corollary 3. *Let* X *be an* n*-dimensional Banach space. For any* $\varepsilon \in (0,1)$, *there exists* Y *a subspace of* X *of dimension* $k \geqslant [(1-\varepsilon)^2 n]$ *such that* $d(Y, l_1^k) \leqslant \frac{\sqrt{n}}{\varepsilon}$.

Using Theorem 3, we obtain an extended factorization result :

Theorem 7. *Let* $X = (\mathbb{R}^n, \|\cdot\|)$ *where* $\|\cdot\|$ *is a norm on* $\mathbb{R}^n$ *such that* B_2^n *is the ellipsoid of minimal volume containing* B_X. *For any* $\varepsilon \in (0,1)$, *there exists* $Y \subset \mathbb{R}^n$ *a subspace of dimension*

$k \geqslant \lfloor (1-\varepsilon)^2 n \rfloor$ such that the following diagram commute

$$\begin{array}{ccc} l_1^k & \xrightarrow{\;i_{1,2}\;} & l_2^k \\ {\scriptstyle \beta}\big\downarrow & \nearrow{\scriptstyle \alpha} & \big\downarrow{\scriptstyle \gamma} \\ (Y, \|\cdot\|) & \xrightarrow[Id_Y]{} & (Y, \|\cdot\|_*), \end{array}$$

where $\|\cdot\|_*$ denotes the dual norm of $\|\cdot\|$. This means that $i_{1,2} = \alpha \circ \beta$ and $Id_Y = \gamma \circ \alpha$. Moreover $\|\beta\| \leqslant 1$, $\|\alpha\| \leqslant \frac{2-\varepsilon}{\varepsilon}$ and $\|\gamma\| \leqslant \frac{2-\varepsilon}{\varepsilon}$.

0.2.3 "Proportional Dvoretzky-Rogers factorization" : nonsymmetric case

The previous results can be stated in terms of symmetric convex bodies since, as it was mentioned previously, it is equivalent to take a norm on $\mathbb{R}^n$ or a symmetric convex body. In [50], Litvak and Tomczak-Jaegermann proved a nonsymmetric version of Theorem 0.10. Precisely, the statement is the following :

Theorem 0.11 (Litvak-Tomczak-Jaegermann). *Let $K \subset \mathbb{R}^n$ be a convex body, such that B_2^n is the ellipsoid of minimal volume containing K. Let $\varepsilon \in (0,1)$ and set $k = [(1-\varepsilon)n]$. There exist vectors $y_1, y_2, ..., y_k$ in K, and an orthogonal projection P in $\mathbb{R}^n$ with rank$(P) \geqslant k$ such that for all scalars $t_1, ..., t_k$*

$$c\varepsilon^3 \left(\sum_{j=1}^k |t_j|^2 \right)^{\frac{1}{2}} \leqslant \left\| \sum_{j=1}^k t_j P y_j \right\|_{PK} \leqslant \frac{6}{\varepsilon} \sum_{j=1}^k |t_j|,$$

where $c > 0$ is a universal constant.

Our aim is, as in the symmetric case, to give a new proof of this result. Introducing this new ingredient given by Theorem 1, we improve the dependence on ε in the previous result and we get :

Theorem 8. *Let $K \subset \mathbb{R}^n$ be a convex body, such that B_2^n is the ellipsoid of minimal volume containing K. For any $\varepsilon \in (0,1)$, there exist $x_1, ..., x_k$ with $k \geqslant (1-\varepsilon)n$ contact points and P an orthogonal projection of rank $\geqslant k$ such that for all $(a_j)_{j \leqslant k}$*

$$\frac{\varepsilon^2}{16} \left(\sum_{j=1}^k |a_j|^2 \right)^{\frac{1}{2}} \leqslant \left\| \sum_{j=1}^k a_j P x_j \right\|_{PK} \leqslant \frac{4}{\varepsilon} \sum_{j=1}^k |a_j|$$

As a direct application of the previous result, we deduce :

Corollary 4. *Let $K \subset \mathbb{R}^n$ be a convex body such that B_2^n is the ellipsoid of minimal volume containing K. For any $\varepsilon \in (0,1)$, there exists P an orthogonal projection of rank $k \geqslant [(1-\varepsilon)n]$ such that*

$$\frac{\varepsilon}{4} B_1^k \subset PK \subset \frac{16}{\varepsilon^2} B_2^k.$$

Moreover, $d(PK, B_1^k) \leqslant \frac{64\sqrt{n}}{\varepsilon^3}$.

By duality, this means that there exists a subspace $E \subset \mathbb{R}^n$ of dimension $k \geqslant [(1-\varepsilon)n]$ such that

$$\frac{\varepsilon^2}{16} B_2^k \subset K \cap E \subset \frac{4}{\varepsilon} B_\infty^k.$$

Moreover, $d(K \cap E, B_\infty^k) \leqslant \frac{64\sqrt{n}}{\varepsilon^3}$.

It is interesting to note that the dependence on the dimension in the previous result is the same as in the symmetric case that we saw in Corollary 3.

0.3 Covariance of random matrices

Random matrices is a vast subject whose applications affect many areas of Mathematics. In this thesis, we investigate the problem of estimating the covariance matrix and more precisely the extension to the matrix setting of some existing vector results.

If X is a random vector in $\mathbb{R}^n$, it is to approximate the covariance matrix of X which is given by $\mathbb{E}XX^t$. By the law of large numbers, we know that when taking many independent copies of X, the average will converge to the covariance matrix of X. More precisely, if $(X_j)_{j \leqslant N}$ are independent copies of X then

$$\frac{1}{N} \sum_{j \leqslant N} X_j X_j^t \underset{N \to \infty}{\longrightarrow} \mathbb{E}XX^t \quad \text{a.s}$$

We are interested in quantifying the convergence i.e finding the minimal number of copies needed in order to approximate the covariance matrix of X. This problem can be formulated as follows :

For $\varepsilon > 0$, find the minimal number $N = N(n, \varepsilon)$ of independent copies of X such that

$$\left\| \frac{1}{N} \sum_{j \leqslant N} X_j X_j^t - \mathbb{E}XX^t \right\| \leqslant \varepsilon,$$

with high probability or even in expectation.

This reminds us the problem of approximating the identity studied by Batson-Spielman-Srivastava [13]. Indeed, let $Id = \sum_{j\leqslant m} x_j x_j^t$ be an identity decomposition on $\mathbb{R}^n$ and $\varepsilon > 0$. Define X a random vector in $\mathbb{R}^n$ by

$$X = \frac{\sqrt{n}}{\|x_j\|_2} x_j \quad \text{with probability } p_j = \frac{\|x_j\|_2^2}{n}.$$

It is easy to see that $\sum_{j\leqslant m} p_j = 1$ and that

$$\mathbb{E}XX^t = \sum_{j\leqslant} p_j \frac{n}{\|x_j\|_2^2} x_j x_j^t = Id.$$

Suppose that we know how to estimate the covariance matrix of X i.e taking $X_1, ..., X_N$ independent copies of X we have

$$\left\| \frac{1}{N} \sum_{j\leqslant N} X_j X_j^t - Id \right\| \leqslant \varepsilon,$$

with positive probability. This implies the existence of $\sigma \subset \{1, ..., m\}$ of size N such that

$$\left\| \sum_{j\leqslant m} s_j x_j x_j^t - Id \right\| \leqslant \varepsilon \quad \text{and} \quad \{j, s_j \neq 0\} = \sigma.$$

If N was of order n, we would recover Theorem 0.1. However, following this procedure, Rudelson [67] obtains N of order $n\log(n)$ which is optimal. Nevertheless, this shows the link between the problem of approximating an identity decomposition and that of approximating the covariance matrix of a random vector.

Note that it is easy to approximate the covariance matrix of a gaussian vector using a number of independent copies proportional to the dimension ; indeed, gaussian vectors satisfy concentration properties with sub-gaussian decay, which allows to use the standard argument of discretization of the unit sphere in order to estimate

$$\sup_{x\in S^{n-1}} \left| \sum_{j\leqslant N} \langle G_j, x\rangle^2 - 1 \right|,$$

where G_j denote independent standard gaussian vectors. After a series of works devoted to this problem ([4], [5],[3]), Adamczack et al. showed that for a random vector satisfying concentration properties with sub-exponential decay, a number of independent copies proportional to the dimension is sufficient to estimate its covariance matrix. We refer to [4] for further information. In this thesis, we were interested in a result due to Srivastava-Vershynin [79] and dealing with the problem of approximating the covariance matrix.

Theorem 0.12 (Srivastava-Vershynin)**.** *Let X_i be independent isotropic random vectors in $\mathbb{R}^n$. Suppose that X_i satisfy a regularity assumption (SR) : there exist $C, \eta > 0$ such that for any orthogonal projection P on $\mathbb{R}^n$*

$$\mathbb{P}\left\{\|PX_i\|_2^2 > t\right\} \leqslant Ct^{-1-\eta} \quad \forall t > C\mathrm{rank}(P). \tag{5}$$

Then, for any $\varepsilon \in (0,1)$ and for

$$N \geqslant C\,\varepsilon^{-2-2/\eta} \cdot n$$

we have

$$\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N} X_iX_i^t - Id\right\| \leqslant \varepsilon. \tag{6}$$

The proof of this theorem consists on randomizing the method of Batson-Spielman-Srivastava [13]. This result has the advantage of covering a large class of distributions ; log-concave vectors as well as random vectors having independent entries with p-moments, where $p > 2$, satisfy (5). However, the covariance estimate is only in expectation.

Our aim is to extend this result to a matrix setting, in the sense of replacing the random vector X by a random matrix of arbitrary size. Note that several studies were devoted to extend known results for random variables/vectors in a matrix framework (see for example [6], [54] and [88])

0.3.1 Estimating the covariance of random matrices

As we have mentioned before, we are interested in extending Theorem 0.12 to a matrix setting i.e given an $n \times m$ random matrix A, we want to approximate $\mathbb{E}AA^t$. The interesting case is when the columns of A are not independent ; otherwise, write $AA^t = \sum_{j\leqslant m} C_jC_j^t$, where $(C_j)_{j\leqslant m}$ are the columns of A, then if each column satisfies (5) then applying Theorem 0.12 to each column and using the triangle inequality, the approximation of $\mathbb{E}AA^t$ follows. Our problem can be formulated in a more general way :

Let B be an $n \times n$ positive semidefinite random matrix satisfying some regularity assumptions. Find the minimal number N of independent copies of B such that $\mathbb{E}\left\|\frac{1}{N}\sum_{i\leqslant N} B_i - \mathbb{E}B\right\|$

is small.

If we suppose that $\mathbb{E}B = Id$ and $\|B\| \leqslant n$ almost surely, then $cn\log(n)$ independent copies of B are sufficient to achieve the approximation. Indeed, take $B_1, ..., B_N, B_1', ..., B_N'$ independent copies of B and $\varepsilon_1, ..., \varepsilon_N$ independent ± 1 Bernouilli variables and write

$$\begin{aligned}
\alpha := \mathbb{E}\left\|\frac{1}{N}\sum_{i\leqslant N} B_i - Id\right\| &= \mathbb{E}\left\|\frac{1}{N}\sum_{i\leqslant N}(B_i - \mathbb{E}B_i')\right\| \\
&\leqslant \mathbb{E}\left\|\frac{1}{N}\sum_{i\leqslant N} B_i - B_i'\right\| \quad \text{by Jensen} \\
&\leqslant 2\mathbb{E}\left\|\frac{1}{N}\sum_{i\leqslant N}\varepsilon_i B_i\right\| \quad \text{by symmetrization} \\
&\leqslant \frac{2}{N}\left(\mathbb{E}\left\|\sum_{i\leqslant N}\varepsilon_i B_i\right\|^p\right)^{\frac{1}{p}} \qquad \text{for any } p \geqslant 1
\end{aligned}$$

Using Khintchine's inequality in the non-commutative setting ([52], [53], see also [57]), we have

$$\left(\mathbb{E}_\varepsilon\left\|\sum_{i\leqslant N}\varepsilon_i B_i\right\|^p\right)^{\frac{1}{p}} \leqslant (c\sqrt{\log(n)} + \sqrt{p})\left\|\sum_{i\leqslant N} B_i^2\right\|^{\frac{1}{2}}$$

Since the B_i's are positive semidefinite then

$$\left\|\sum_{i\leqslant N} B_i^2\right\|^{\frac{1}{2}} \leqslant \max_{i\leqslant N}\|B_i\|^{\frac{1}{2}} \cdot \left\|\sum_{i\leqslant N} B_i\right\|$$

Thus if $\|B\| \leqslant n$ almost surely, we get

$$\begin{aligned}
\alpha &\leqslant \frac{c\sqrt{n\log(n)}}{N}\cdot\mathbb{E}\left\|\sum_{i\leqslant N} B_i\right\|^{\frac{1}{2}} \\
&\leqslant \sqrt{\frac{n\log(n)}{N}}\cdot\left[\mathbb{E}\left\|\sum_{i\leqslant N}\frac{1}{N}B_i\right\|\right]^{\frac{1}{2}} \\
&\leqslant c\sqrt{\frac{n\log(n)}{N}}\cdot\left(\sqrt{\alpha}+1\right)
\end{aligned}$$

By a simple calculation, we deduce that

$$\alpha = \mathbb{E}\left\|\frac{1}{N}\sum_{i\leqslant N} B_i - Id\right\| \leqslant c\sqrt{\frac{n\log(n)}{N}}$$

and therefore it is sufficient to take N of order $cn\log(n)$ in order to make α small.

Without regularity assumptions, nothing better can be said ; indeed, taking B uniformly distributed over $\{ne_ie_i^t\}_{i\leqslant n}$, where $(e_i)_{i\leqslant n}$ denotes the canonial basis of $\mathbb{R}^n$, one can see that $cn\log(n)$ copies are needed.

De Carli Silva-Harvey-Sato [26] extended the result of Batson-Spielman-Srivastava [13] to a matrix setting. By randomizing the method of De Carli Silva-Harvey-Sato, we obtain the following :

Theorem 9. *Let B be an $n\times n$ positive semidefinite random matrix. Suppose that $\mathbb{E}B = Id$ and B satisfies a regularity assumption (MSR) : there exist $c, \eta > 0$ such that for any orthogonal projection P on $\mathbb{R}^n$,*

$$\mathbb{P}\left\{\|PBP\| > t\right\} \leqslant ct^{-1-\eta} \quad \forall t > c\,\mathrm{rank}(P). \tag{7}$$

Then, for any $\varepsilon \in (0,1)$ and any

$$N \geqslant c\,\varepsilon^{-2-2/\eta}\cdot n$$

we have

$$\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N} B_i - Id\right\| \leqslant \varepsilon, \tag{8}$$

where $B_1, ..., B_N$ are independent copies of B.

This generalizes Theorem 0.12 ; indeed, let X be an isotropic random vector in $\mathbb{R}^n$ satisfying (5). Take $B = XX^t$, then $\mathbb{E}B = Id$ and since $\|PBP\| = \|PX\|_2^2$ then B satisfies (7). Therefore, applying Theorem 9 to B, we get (6).

In order to prove Theorem 9, we establish estimates on the smallest and largest eigenvalue of the sum of random matrices satisfying (MSR).

In [9], Bai-Yin showed that if X is a random vector in $\mathbb{R}^n$ with centered i.i.d entries of variance 1 and with finite fourth moment, then if $N = N(n)$ such that $h = \lim\limits_{n\to\infty}\frac{n}{N} \in (0,1)$ we have

$$\lambda_{min}\left(\frac{1}{N}\sum_{j\leqslant N} X_jX_j^t\right) \underset{n\to\infty}{\longrightarrow} (1-\sqrt{h})^2 \quad a.s$$

et

$$\lambda_{max}\left(\frac{1}{N}\sum_{j\leqslant N}X_jX_j^t\right)\underset{n\to\infty}{\longrightarrow}(1+\sqrt{h})^2 \quad a.s$$

The estimates of the smallest and largest eigenvalue of the sum of random matrices satisfying (MSR), given by Theorem 9, can be seen as non-asymptotic matrix version of the result of Bai-Yin.

We give examples where our result applies. The case of log-concave matrices is among these ones and will be detailed in the next section.

0.3.2 Log-concave matrices

A probability measure μ on $\mathbb{R}^n$ is called log-concave if for any $0 < t < 1$ and any $A, B \subset \mathbb{R}^n$ compact sets of positive measure, we have

$$\mu\left((1-t)A + tB\right) \geqslant \mu(A)^{1-t}\mu(B)^t.$$

Borell ([17],[16]) gave a characterization of log-concave measures : μ is a log-concave measure on $\mathbb{R}^n$ if and only if the density f of μ with respect to the Lebesgue measure is log-concave i.e $\log(f)$ is a concave function. Finally, a random vector in $\mathbb{R}^n$ is called log-concave if its distribution is log-concave.

Log-concave measures play an important role in the geometry of convex bodies ; indeed, if K is a convex body then $\mathbf{1}_K$ is a log-concave measure. Conversely, given a log-concave measure, Ball [10] associate to this measure a convex body. Many work on log-concave measures have been made in recent years ; one of the important results is that of Paouris ([58], [59], [60]) who establishes large deviation inequalities and small ball probability estimates for log-concave vectors. To make the link with the previous section, one of the major problems is to estimate the covariance matrix of a log-concave vector ; this was solved by Adamczack et al. [4]. We are interested in the notion of log-concavity in a matrix framework. Since an $n \times m$ matrix can be seen as a vector in $\mathbb{R}^{nm}$, then one can naturally define the notion of log-concave matrices ; precisely, an $n \times m$ random matrix A is log-concave if its distribution is a log-concave measure on $\mathbb{R}^{nm}$. It remains to define the isotropic condition.

Definition 0.1. *Let A be an $n \times m$ random matrix having $(C_i)_{i\leqslant m}$ as columns. We will say that A is an isotropic log-concave matrix if $A'^t = \sqrt{m}(C_1^t, .., C_m^t)$ is an isotropic log-concave vector in $\mathbb{R}^{nm}$.*

Therefore when A is isotropic, we have $\mathbb{E}AA^t = Id$. It is easy to deduce concentration properties of these matrices using the results of Paouris. We use a more refined result due to Guédon-Milman [36], which states that with high probability, an isotropic log-concave vector in $\mathbb{R}^n$ lies in a thin shell of radius almost $\sqrt{n}$. Therefore, log-concave matrices satisfy the following :

Proposition 5. *Let A be an $n \times m$ isotropic log-concave matrix. If $B = AA^t$, then for any orthogonal projection P on $\mathbb{R}^n$, we have*

$$\mathbb{P}\left\{|\mathrm{Tr}(PB) - \mathrm{rank}(P)| \geqslant t.\mathrm{rank}(P)\right\} \leqslant C\exp\left(-ct^3\sqrt{m.\mathrm{rank}(P)}\right) \quad \forall t \leqslant 1. \tag{9}$$

Using the previous property, it is clear that an isotropic log-concave matrix satisfies (MSR) and thus the results of the previous section can be applied.

The concentration inequality given by Proposition 5 are strong enough to allow us obtain results with high probability rather than in expectation. However, we need the matrix to be sufficiently rectangular. More precisely, we prove the following :

Theorem 10. *Let A be an $n \times m$ isotropic log-concave matrix. For any $\varepsilon \in (0,1)$, if $m \geqslant \frac{C}{\varepsilon^6}\left[\log(CnN)\right]^2$, taking*

$$N \geqslant \frac{96n}{\varepsilon^2}$$

independent copies $A_1, ..., A_N$ of A, then with probability $\geqslant 1 - \exp(-c\varepsilon^3\sqrt{m})$ we have

$$\left\|\frac{1}{N}\sum_{i=1}^{N} A_iA_i^t - Id\right\| \leqslant \varepsilon$$

Note that the number of independent copies needed in the previous result is optimal as it can be seen by taking gaussian matrices. The set of isotropic log-concave matrices we defined, contains a large class of examples. Therefore, we get :

Proposition 6. *Let A be an $n \times m$ random matrix whose density with respect to Lebesgue is given by*

$$G(A) = \exp\left(-f(s_1(A), ..., s_k(A))\right),$$

where f is an absolutely symmetric convex function, properly normalized as above and $k = \min(n,m)$.
Suppose that $m \geqslant \frac{C}{\varepsilon^6}\left[\log(\frac{Cn}{\varepsilon})\right]^2$ and $n \geqslant \frac{C}{\varepsilon^6}\left[\log(\frac{Cm}{\varepsilon})\right]^2$, taking $N = \frac{96\max(n,m)}{\varepsilon^2}$ then with proba-

bility $\geqslant 1 - \exp(-c\varepsilon^3\sqrt{k})$ *we have*

$$1 - \varepsilon \leqslant \lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^{N} A_i A_i^t\right) \leqslant \lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} A_i A_i^t\right) \leqslant 1 + \varepsilon$$

and

$$(1-\varepsilon)\frac{n}{m} \leqslant \lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^{N} A_i^t A_i\right) \leqslant \lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} A_i^t A_i\right) \leqslant (1+\varepsilon)\frac{n}{m}$$

Recently, a new proof of the result of Paouris was given in [2]. In [1], it is shown that the method in [2] extends to the case of convex measures. Therefore, concentration properties for isotropic $(-\frac{1}{r})$-concave vectors were established. Using these properties, we can obtain analogues of the previous results in the case of isotropic $(-\frac{1}{r})$-concave matrices.

0.4 References

This thesis consists of three papers that were all submitted to international journals. Thus, the content of the chapters is essentially the content of these articles although some additional results were added. Our papers can be found on Arxiv :

1. "Restricted invertibility and the banach-mazur distance to the cube", Available at arXiv :1206.0654.
2. "A note on column subset selection", Available at arXiv :1212.0976.
3. "Estimating the covariance of random matrices", Available at arXiv :1301.6607.

Results of chapter 1 and 3 are essentially contained in the first paper. The content of chapter 2 is basically the second paper, and finally, the results of chapters 4 and 5 are contained in the third paper.

Chapitre 1

Restricted invertibility

1.1 Introduction

An $n \times m$ matrix U can be seen as a linear operator from l_2^m into l_2^n. The invertibility of a matrix is a very precious property which means that the corresponding operator is an isomorphism. Of course the important property is the injectivity which implies that the matrix is an isomorphism on its image. Unfortunately, all matrices are not injective especially in the case where the dimension of the domain is larger than the dimension of the codomain. However, one can search for a subspace such that the restriction of the operator on this subspace is injective. Finding a coordinate subspace is much more convenient, since the restriction can be traced inside the matrix as a choice of a block of columns. Let us be more precise and introduce the target of this chapter. Let U be an $n \times m$ matrix, we are interested in two problems :

– **Problem 1** : Extract a "large" number of linearly independent columns of U. Denote $\sigma \subset \{1, ..., m\}$ the set of indices of the columns chosen and U_σ the restriction of U to $\text{span}\{(e_j)_{j\in\sigma}\}$.

– **Problem 2** : Estimate $s_{\min}(U_\sigma)$.

Of course, our aim is to solve the two problems simultaneously to get what is called a restricted invertibility principle. Looking at Problem 1, the "large" number is clearly the definition of the rank of a matrix. However, this will not be convenient since this notion is not stable by perturbation. Take for example a diagonal matrix D of size $n \times n$ with all entries equal δ except one which is equal to 1. Clearly, D is of full rank but is mostly directed in one direction. When δ is small, modifying D by a small perturbation $\delta \cdot Id$, it becomes of rank 1.

Considering the previous remark, we define the notion of stable rank by

$$\mathrm{srank}(U) = \frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2}$$

Clearly we have $\mathrm{srank}(U) \leqslant \mathrm{rank}(U)$. Basicaly, the stable rank will be big if the matrix is well distributed over the eigenspaces and will be small if it is distributed over few eigenspaces. Looking at the previous example with δ sufficiently small, the stable rank of D is almost one before and after perturbation. From now on, the "large" number we are searching for will be the stable rank. So our problem reduces to the following :

Given an $n \times m$ matrix U, find a subset σ of the columns of U of size almost the stable rank and estimate the smallest singular value of U_σ.

By estimating the smallest singular value, we mean giving a lower bound for it. Note that if $s_{|\sigma|}(U_\sigma)$ is bounded away from zero, then clearly U_σ is of rank $|\sigma|$ and therefore injective. Let us also note that estimating the smallest singular value is equivalent to estimate the norm of the inverse of the restricted matrix U_σ. Indeed, for $x \in \mathbb{R}^{|\sigma|}$ write

$$\begin{aligned}
\|U_\sigma x\|_2^2 &= \langle U_\sigma x, U_\sigma x\rangle \\
&= \langle U_\sigma^* U_\sigma x, x\rangle \\
&\geqslant s_{\min}^2(U_\sigma)\|x\|_2^2
\end{aligned}$$

Therefore from now on, when estimating the norm of the inverse or estimating the smallest singular value, we will be dealing with the same problem. The restricted invertibility principle was first studied by Bourgain-Tzafriri [19] who proved a result for square matrices with columns of norm 1.

Theorem 1.1 (Bourgain-Tzafriri)**.** *Let T be an $n \times n$ matrix satisfying $\|Te_j\|_2 = 1$, where $(e_j)_{j\leqslant n}$ denotes the canonical basis of $\mathbb{R}^n$. Then there exists $\sigma \subset \{1, .., n\}$ with*

$$|\sigma| \geqslant \left\lfloor c\frac{n}{\|T\|^2} \right\rfloor,$$

such that for any choice of scalars $(a_j)_{j\in\sigma}$

$$\left\| \sum_{j\in\sigma} a_j T e_j \right\|_2 \geqslant d \left(\sum_{j\in\sigma} a_j^2 \right)^{\frac{1}{2}},$$

where c and d are universal constants.

Note that since $\|Te_j\|_2 = 1$ then $\|T\|_{HS}^2 = n$ so that the size of the extracted matrix is proportional to the stable rank of T. The conclusion of the theorem also means that $s_{\min}(T_\sigma) \geqslant d$. This result is usually known as the Bourgain-Tzafriri restricted invertibility principle. The proof of this theorem [19] uses the selectors (a probabilistic argument), Sauer-Shelah lemma ([70],[73]) which is a combinatorial argument and Grothendieck's factorization theorem (see [27] and [61]). In [87], Tropp gave a randomized algorithm to realize Grothendieck's factorization theorem and therefore he was able to give a randomized algorithm for the Bourgain-Tzafriri restricted invertibility principle.

Back to Bourgain-Tzafriri's result, one can interpret it in another way. Since the identity operator can be decomposed in the form $Id = \sum_{j\leqslant n} e_j e_j^t$ where $(e_j)_{j\leqslant n}$ is the canonical basis of $\mathbb{R}^n$, then Theorem 1.1 states that one can find a large part of this basis (of cardinality greater than $c\frac{n}{\|T\|^2}$), on the span of which, the operator T is invertible and the norm of its inverse is controlled by an absolute constant.

Looking at this formulation, Vershynin [90] generalized this result for any decomposition of the identity and improved the estimate for the size of the subset. Vershynin proved also a non trivial upper bound which will not be our target for now. For this reason, we will just state a part of Vershynin's result, the one involved with the restricted invertibility principle, while the full result will be the subject of the next chapter and a full statement can be found there. Using a technical iteration scheme based on Theorem 1.1, combined with a theorem of Kashin-Tzafriri [43] which we will discuss later in this chapter, Vershynin obtained the following :

Theorem 1.2 (Vershynin)**.** *Let $Id = \sum_{j\leqslant m} x_j x_j^t$ and let T be a linear operator on l_2^n. For any $\varepsilon \in (0,1)$, one can find $\sigma \subset \{1,\dots,m\}$ with*

$$|\sigma| \geqslant \lfloor (1-\varepsilon)\mathrm{srank}(T) \rfloor = \left\lfloor (1-\varepsilon)\frac{\|T\|_{\mathrm{HS}}^2}{\|T\|^2} \right\rfloor$$

such that

$$\left\| \sum_{j\in\sigma} a_j \frac{Tx_j}{\|Tx_j\|_2} \right\|_2 \geqslant c(\varepsilon) \left(\sum_{j\in\sigma} a_j^2 \right)^{\frac{1}{2}}$$

for all scalars $(a_j)_{j\in\sigma}$.

This clearly generalizes the Bourgain-Tzafriri restricted invertibility principle. Indeed, if $\|Te_j\|_2 = 1$ then $\mathrm{srank}(T) = \frac{n}{\|T\|^2}$ and Vershynin's result states that there exists $\sigma \subset \{1,\dots,n\}$

of size $(1-\varepsilon)\frac{n}{\|T\|^2}$ such that

$$\left\|\sum_{j\in\sigma} a_j T e_j\right\|_2 \geqslant c(\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}$$

for all scalars $(a_j)_{j\in\sigma}$. This recovers Theorem 1.1 with the possibility of finding a subset of size almost the stable rank of T while in Theorem 1.1, we couldn't go further than a proportion of the stable rank. The constant $c(\varepsilon)$ plays a crucial role in applications, however the estimate of Vershynin was not of a good order. This will be more detailed in the next chapter. Finally, let us note that having these normalizing factors $\|Tx_j\|_2$ in Vershynin's result is convenient for applications as we may see later.

Back to the original restricted invertibility problem, a recent work of Spielman-Srivastava [74] provides the best known estimate for the norm of the inverse matrix. Their proof gives a deterministic algorithm based on the method invented by Batson-Spielman-Srivastava [13] and makes use of linear algebra tools, while the previous works on the subject employed probabilistic, combinatorial and functional-analytic arguments.

More precisely, Spielman-Srivastava proved the following :

Theorem 1.3 (Spielman-Srivastava)**.** *Let $x_1,\dots x_m \in \mathbb{R}^n$ such that $Id = \sum_i x_i x_i^t$ and let $0<\varepsilon<1$. For every linear operator $T : \ell_2^n \to \ell_2^n$ there exists a subset $\sigma \subset \{1,\dots,m\}$ of size $|\sigma| \geqslant \left\lfloor (1-\varepsilon)^2 \frac{\|T\|_{\mathrm{HS}}^2}{\|T\|^2}\right\rfloor$ for which $\{Tx_i\}_{i\in\sigma}$ is linearly independent and*

$$\lambda_{\min}\left(\sum_{i\in\sigma}(Tx_i)(Tx_i)^t\right) > \frac{\varepsilon^2\|T\|_{\mathrm{HS}}^2}{m},$$

where $\lambda_{\min}$ is computed on $\mathrm{span}\{Tx_i\}_{i\in\sigma}$ or simply here $\lambda_{\min}$ denotes the smallest nonzero eigenvalue of the corresponding operator.

Equivalently, for all scalars $(a_j)_{j\in\sigma}$

$$\left\|\sum_{j\in\sigma} a_j T x_j\right\|_2 \geqslant \varepsilon\frac{\|T\|_{\mathrm{HS}}}{\sqrt{m}}\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}.$$

One can view the previous result as a restricted invertibility principle for rectangular matrices. Given, as above, an identity decomposition and a linear operator T, one can associate to these an $n\times m$ matrix U whose columns are the vectors $(Tx_j)_{j\leqslant m}$. Since $Id = \sum_j x_j x_j^t$, one can easily check that

$$U \cdot U^t = T \cdot T^t = \sum_{j \leqslant m} (Tx_j) \cdot (Tx_j)^t.$$

Hence,

$$\|U\|_{\mathrm{HS}} = \|T\|_{\mathrm{HS}} \quad \text{and} \quad \|U\| = \|T\|,$$

and thus the previous result can be written in terms of the rectangular matrix U. Beside producing a deterministic algorithm, Spielman-Srivastava's result gives the best known dependence on ε. However, their result is without the normalizing factors appearing in Vershynin's one and which are crucial for our study to the Banach-Mazur distance to the cube in chapter 3. On one hand, Vershynin's result has these normalizing factors but produces a bad dependence on ε and on the other hand Spielman-Srivastava's result gives a good dependence on ε but without these normalizing factors. It becomes natural to search for a result which interpolates the two. We will be able to prove the restricted invertibility principle for any rectangular matrix, with any normalizing factors for the columns and with a good dependence on ε. Before giving the precise statement, let us introduce some notations :

If D is an $m \times m$ diagonal matrix with diagonal entries $(\alpha_j)_{j \leqslant m}$, we set $I_D := \{j \leqslant m \mid \alpha_j \neq 0\}$ and write D_σ^{-1} for the restricted inverse of D i.e the diagonal matrix whose diagonal entries are the inverse of the respective entries of D for indices in σ and zero elsewhere. The main result of this chapter is the following :

Theorem 1.4. *Given an $n \times m$ matrix U and a diagonal $m \times m$ matrix D with $(\alpha_j)_{j \leqslant m}$ on its diagonal, with the property that* $\mathrm{Ker}(D) \subset \mathrm{Ker}(U)$*, then for any $\varepsilon \in (0,1)$ there exists $\sigma \subset I_D$ with*

$$|\sigma| \geqslant \left\lfloor (1-\varepsilon)^2 \mathrm{srank}(U) \right\rfloor = \left\lfloor (1-\varepsilon)^2 \frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2} \right\rfloor$$

such that

$$s_{\min}\left(U_\sigma D_\sigma^{-1}\right) > \frac{\varepsilon \|U\|_{\mathrm{HS}}}{\|D\|_{\mathrm{HS}}},$$

where $s_{\min}$ denotes the smallest singular value.

Equivalently, for all scalars $(a_j)_{j \in \sigma}$

$$\left\| \sum_{j \in \sigma} a_j \frac{Ue_j}{\alpha_j} \right\|_2 \geqslant \varepsilon \frac{\|U\|_{\mathrm{HS}}}{\|D\|_{\mathrm{HS}}} \left(\sum_{j \in \sigma} a_j^2 \right)^{\frac{1}{2}}.$$

Let $Id = \sum_{j \leqslant m} x_j x_j^t$ be an identity decomposition and T a linear operator on l_2^n. Define U the $n \times m$ matrix whose columns are the $(Tx_j)_{j \leqslant m}$. It is easy to check that $TT^t = UU^t$ so that srank(U) = srank(T).

– If $D = Id$, then clearly $\text{Ker}(D) = \{0\} \subset \text{Ker}(U)$. Apply Theorem 1.4 with U and D to find σ of size

$$|\sigma| \geqslant \left\lfloor (1-\varepsilon)^2 \text{srank}(U) \right\rfloor = \left\lfloor (1-\varepsilon)^2 \text{srank}(T) \right\rfloor$$

such that

$$\left\| \sum_{j\in\sigma} a_j T x_j \right\|_2 = \left\| \sum_{j\in\sigma} a_j U e_j \right\|_2 \geqslant \varepsilon \frac{\|U\|_{\text{HS}}}{\|D\|_{\text{HS}}} \left(\sum_{j\in\sigma} a_j^2 \right)^{\frac{1}{2}} = \varepsilon \frac{\|T\|_{\text{HS}}}{\sqrt{m}} \left(\sum_{j\in\sigma} a_j^2 \right)^{\frac{1}{2}}.$$

This means that Theorem 1.4 implies Theorem 1.3.

– If $D = \text{diag}\,(\alpha_1, ..., \alpha_m)$ with $\alpha_j = \|Tx_j\|_2$ then it is easy to see that $\text{Ker}(D) \subset \text{Ker}(U)$ and $\|D\|_{\text{HS}} = \|U\|_{\text{HS}}$. Apply Theorem 1.4 with U and D to find σ of size

$$|\sigma| \geqslant \left\lfloor (1-\varepsilon)^2 \text{srank}(U) \right\rfloor = \left\lfloor (1-\varepsilon)^2 \text{srank}(T) \right\rfloor$$

such that

$$\left\| \sum_{j\in\sigma} a_j \frac{T x_j}{\|Tx_j\|_2} \right\|_2 = \left\| \sum_{j\in\sigma} a_j \frac{U e_j}{\alpha_j} \right\|_2 \geqslant \varepsilon \frac{\|U\|_{\text{HS}}}{\|D\|_{\text{HS}}} \left(\sum_{j\in\sigma} a_j^2 \right)^{\frac{1}{2}} = \varepsilon \left(\sum_{j\in\sigma} a_j^2 \right)^{\frac{1}{2}}$$

This means that Theorem 1.4 implies Theorem 1.2.

1.2 Proof of Theorem 1.4

Since the rank and the eigenvalues of $(U_\sigma D_\sigma^{-1})^t \cdot (U_\sigma D_\sigma^{-1})$ and $(U_\sigma D_\sigma^{-1}) \cdot (U_\sigma D_\sigma^{-1})^t$ are the same, it suffices to prove that $(U_\sigma D_\sigma^{-1}) \cdot (U_\sigma D_\sigma^{-1})^t$ has rank equal to $k = |\sigma|$ and its smallest positive eigenvalue is greater than $\varepsilon^2 \frac{\|U\|_{\text{HS}}^2}{\|D\|_{\text{HS}}^2}$. Note that

$$(U_\sigma D_\sigma^{-1}) \cdot (U_\sigma D_\sigma^{-1})^t = \sum_{j\in\sigma} \left(U D_\sigma^{-1} e_j \right) \cdot \left(U D_\sigma^{-1} e_j \right)^t = \sum_{j\in\sigma} \left(\frac{U e_j}{\alpha_j} \right) \cdot \left(\frac{U e_j}{\alpha_j} \right)^t$$

We are going to construct the matrix $A_k = \sum_{j\in\sigma} (U D_\sigma^{-1} e_j) \cdot (U D_\sigma^{-1} e_j)^t$ by iteration. We start by setting $A_0 = 0$ and at each step we will be adding a rank one matrix $\left(\frac{Ue_j}{\alpha_j}\right) \cdot \left(\frac{Ue_j}{\alpha_j}\right)^t$ for a suitable j, which will give a new positive eigenvalue. This will guarantee that the vector $U D_\sigma^{-1} e_j$ chosen in each step is linearly independent from the previous ones.

If A and B are symmetric matrices, we write $A \preceq B$ if $B - A$ is a positive semidefinite matrix. Recall the Sherman-Morrison Formula which will be needed in the proof. For any invertible matrix A and any vector v we have

$$(A + v \cdot v^t)^{-1} = A^{-1} - \frac{A^{-1}v \cdot v^t A^{-1}}{1 + v^t A^{-1} v}.$$

We will also use the following lemma which appears as Lemma 6.3 in [78] :

Lemma 1.5. *Suppose that $A \succeq 0$ has q nonzero eigenvalues, all greater than $b' > 0$. If $v \neq 0$ and*

$$v^t (A - b'I)^{-1} v < -1, \tag{1.1}$$

then $A + vv^t$ has $q+1$ nonzero eigenvalues, all greater than b'.

Proof. The proof of the lemma is simple and makes use of the Sherman-Morrison formula. Denote $\lambda_1 \geqslant \lambda_2 \geqslant .. \geqslant \lambda_q > b'$ the eigenvalues of A and $\lambda'_1 \geqslant .. \geqslant \lambda'_q \geqslant \lambda'_{q+1} \geqslant 0$ the eigenvalues of $A + vv^t$. By the Cauchy interlacing Theorem we have

$$\lambda'_1 \geqslant \lambda_1 \geqslant \lambda'_2 \geqslant .. \geqslant \lambda'_q \geqslant \lambda_q \geqslant \lambda'_{q+1}$$

Since $\lambda_q > b'$ then $\lambda'_q > b'$ and it remains to prove that $\lambda'_{q+1} > b'$. Now write

$$\mathrm{Tr}(A + vv^t - b'I)^{-1} = \sum_{j \leqslant q+1} \frac{1}{\lambda'_j - b'} - \sum_{j > q+1} \frac{1}{b'},$$

and

$$\mathrm{Tr}(A - b'I)^{-1} = \sum_{j \leqslant q} \frac{1}{\lambda_j - b'} - \sum_{j > q} \frac{1}{b'}$$

then

$$\begin{aligned} \mathrm{Tr}(A + vv^t - b'I)^{-1} - \mathrm{Tr}(A - b'I)^{-1} &= \sum_{j \leqslant q} \left[\frac{1}{\lambda'_j - b'} - \frac{1}{\lambda_j - b'} \right] + \frac{1}{\lambda'_{q+1} - b'} - \frac{1}{b'} \\ &\leqslant \frac{1}{\lambda'_{q+1} - b'} - \frac{1}{b'} \end{aligned}$$

By the Sherman-Morrison's formula we have :

$$(A + vv^t - b'I)^{-1} = (A - b'I)^{-1} - \frac{(A - b'I)^{-1} vv^t (A - b'I)^{-1}}{1 + v^t (A - b'I)^{-1} v}$$

Now taking the trace we get :

$$\mathrm{Tr}(A + vv^t - b'I)^{-1} - \mathrm{Tr}(A - b'I)^{-1} = -\frac{v^t(A - b'I)^{-2}v}{1 + v^t(A - b'I)^{-1}v}$$

Since $(A - b'I)^{-2}$ is positive definite then using the hypothesis (1.1), one can see that the right hand side in the previous equality is positive. Therefore we get :

$$\frac{1}{\lambda'_{q+1} - b'} - \frac{1}{b'} > 0$$

wich means that $\lambda'_{q+1} > b'$.

□

For any symmetric matrix A and any $b > 0$, we define

$$\phi(A, b) = \mathrm{Tr}\left(U^t(A - bI)^{-1}U\right)$$

as the potential corresponding to the barrier b.
At each step l, the matrix already constructed is denoted by A_l and the barrier by b_l. Suppose that A_l has l nonzero eigenvalues all greater than b_l. As mentioned before, we will try to construct A_{l+1} by adding a rank one matrix $v \cdot v^t$ to A_l so that A_{l+1} has $l+1$ nonzero eigenvalues all greater than $b_{l+1} = b_l - \delta$, with $\delta > 0$, and $\phi(A_{l+1}, b_{l+1}) \leqslant \phi(A_l, b_l)$. Note that

$$\begin{aligned}\phi(A_{l+1}, b_{l+1}) &= \mathrm{Tr}\left(U^t(A_l + vv^t - b_{l+1}I)^{-1}U\right)\\ &= \mathrm{Tr}\left(U^t(A_l - b_{l+1}I)^{-1}U\right) - \mathrm{Tr}\left(\frac{U^t(A_l - b_{l+1}I)^{-1}vv^t(A_l - b_{l+1}I)^{-1}U}{1 + v^t(A_l - b_{l+1}I)^{-1}v}\right)\\ &= \phi(A_l, b_{l+1}) - \frac{v^t(A_l - b_{l+1}I)^{-1}UU^t(A_l - b_{l+1}I)^{-1}v}{1 + v^t(A_l - b_{l+1}I)^{-1}v}.\end{aligned}$$

So, in order to have $\phi(A_{l+1}, b_{l+1}) \leqslant \phi(A_l, b_l)$, we must choose a vector v verifying

$$-\frac{v^t(A_l - b_{l+1}I)^{-1}UU^t(A_l - b_{l+1}I)^{-1}v}{1 + v^t(A_l - b_{l+1}I)^{-1}v} \leqslant \phi(A_l, b_l) - \phi(A_l, b_{l+1}). \tag{1.2}$$

Since $v^t(A_l - b_{l+1}I)^{-1}UU^t(A_l - b_{l+1}I)^{-1}v$ and $(\phi(A_l, b_l) - \phi(A_l, b_{l+1}))$ are positive, choosing v verifying conditions (1.1) and (1.2) is equivalent to choosing v which satisfies the following :

$$v^t(A_l - b_{l+1}I)^{-1}UU^t(A_l - b_{l+1}I)^{-1}v \leqslant (\phi(A_l, b_l) - \phi(A_l, b_{l+1}))\left(-1 - v^t(A_l - b_{l+1}I)^{-1}v\right)$$

Since $UU^t \preceq \|U\|^2 Id$ and $(A_l - b_{l+1}I)^{-1}$ is symmetric, it is sufficient to choose v such that

$$v^t(A_l - b_{l+1}I)^{-2}v \leqslant \frac{1}{\|U\|^2}\left(\phi(A_l, b_l) - \phi(A_l, b_{l+1})\right)\left(-1 - v^t(A_l - b_{l+1}I)^{-1}v\right) \tag{1.3}$$

Recall the notation $I_D := \{j \leqslant m \mid \alpha_j \neq 0\}$ where $(\alpha_j)_{j\leqslant m}$ are the diagonal entries of D. Since we assumed that $\text{Ker}(D) \subset \text{Ker}(U)$, we have

$$\|U\|_{\text{HS}}^2 = \sum_{j\leqslant m} \|Ue_j\|_2^2 = \sum_{j\in I_D} \|Ue_j\|_2^2 \leqslant |I_D| \cdot \|U\|^2,$$

and thus $|I_D| \geqslant \frac{\|U\|_{\text{HS}}^2}{\|U\|^2}$. At each step, we will select a vector v satisfying (1.3) among $(\frac{Ue_j}{\alpha_j})_{j\in I_D}$. Therefore, our task is to find $j \in I_D$ such that

$$(Ue_j)^t(A_l - b_{l+1}I)^{-2}Ue_j \leqslant \frac{\phi(A_l, b_l) - \phi(A_l, b_{l+1})}{\|U\|^2}\left(-\alpha_j^2 - (Ue_j)^t(A_l - b_{l+1}I)^{-1}Ue_j\right) \tag{1.4}$$

The existence of such a $j \in I_D$ is guaranteed by the fact that condition (1.4) holds true if we take the sum over all $(\frac{Ue_j}{\alpha_j})_{j\in D}$. The hypothesis $\text{Ker}(D) \subset \text{Ker}(U)$ implies that :

- $\sum_{j\in I_D} (Ue_j)^t(A_l - b_{l+1}I)^{-2}Ue_j = \text{Tr}\left(U^t(A_l - b_{l+1}I)^{-2}U\right)$,
- $\sum_{j\in I_D} (Ue_j)^t(A_l - b_{l+1}I)^{-1}Ue_j = \text{Tr}\left(U^t(A_l - b_{l+1}I)^{-1}U\right)$.

Therefore it is enough to prove that, at each step, one has

$$\text{Tr}(U^t(A_l - b_{l+1}I)^{-2}U) \leqslant \frac{\phi(A_l, b_l) - \phi(A_l, b_{l+1})}{\|U\|^2}\left(-\|D\|_{\text{HS}}^2 - \phi(A_l, b_{l+1})\right) \tag{1.5}$$

The next lemma will determine the conditions required at each step in order to prove (1.5).

Lemma 1.6. *Suppose that A_l has l nonzero eigenvalues all greater than b_l, and write Z for the orthogonal projection onto the kernel of A_l. If*

$$\phi(A_l, b_l) \leqslant -\|D\|_{\text{HS}}^2 - \frac{\|U\|^2}{\delta} \tag{1.6}$$

and

$$0 < \delta < b_l \leqslant \delta\frac{\|ZU\|_{\text{HS}}^2}{\|U\|^2}, \tag{1.7}$$

then there exists $i \in I_D$ such that $A_{l+1} := A_l + \left(\frac{Ue_i}{\alpha_i}\right)\cdot\left(\frac{Ue_i}{\alpha_i}\right)^t$ has $l+1$ nonzero eigenvalues all greater than $b_{l+1} := b_l - \delta$ and $\phi(A_{l+1}, b_{l+1}) \leqslant \phi(A_l, b_l)$.

Proof. As mentioned before, it is enough to prove inequality (1.5). We set $\Delta_l := \phi(A_l, b_l) - \phi(A_l, b_{l+1})$. By (1.6), we get

$$\phi(A_l, b_{l+1}) \leqslant -\|D\|_{\mathrm{HS}}^2 - \frac{\|U\|^2}{\delta} - \Delta_l.$$

Inserting this in (1.5), we see that it is sufficient to prove the following inequality :

$$\mathrm{Tr}\left(U^t(A_l - b_{l+1}I)^{-2}U\right) \leqslant \Delta_l \left(\frac{\Delta_l}{\|U\|^2} + \frac{1}{\delta}\right). \tag{1.8}$$

Now, denote by P the orthogonal projection onto the image of A_l. We set

$$\phi^P(A_l, b_l) := \mathrm{Tr}\left(U^t P(A_l - b_l I)^{-1} PU\right) \quad \text{and} \quad \Delta_l^P := \phi^P(A_l, b_l) - \phi^P(A_l, b_{l+1})$$

and use similar notation for Z. Since P, Z and A_l commute, one can write

$$\Delta_l = \Delta_l^P + \Delta_l^Z \quad \text{and} \quad \phi(A_l, b_l) = \phi^P(A_l, b_l) + \phi^Z(A_l, b_l).$$

Note that

$$\begin{aligned}(A_l - b_l I)^{-1} - (A_l - b_{l+1}I)^{-1} &= (A_l - b_l I)^{-1}(b_l I - A_l + A_l - b_{l+1}I)(A_l - b_{l+1}I)^{-1} \\ &= \delta(A_l - b_l I)^{-1}(A_l - b_{l+1}I)^{-1}\end{aligned}$$

and since $P(A_l - b_l I)^{-1}P$ and $P(A_l - b_{l+1}I)^{-1}P$ are positive semidefinite, we have

$$U^t P(A_l - b_l I)^{-1} PU - U^t P(A_l - b_{l+1}I)^{-1} PU \succeq \delta U^t P(A_l - b_{l+1}I)^{-2} PU.$$

Inserting this in (1.8), it is enough to prove that :

$$\mathrm{Tr}\left(U^t Z(A_l - b_{l+1}I)^{-2} ZU\right) \leqslant \Delta_l \left(\frac{\Delta_l}{\|U\|^2} + \frac{1}{\delta}\right) - \frac{\Delta_l^P}{\delta}.$$

Since $A_l Z = 0$, we have

$$\mathrm{Tr}(U^t Z(A_l - b_{l+1}I)^{-2} ZU) = \frac{\|ZU\|_{\mathrm{HS}}^2}{b_{l+1}^2}$$

and

$$\Delta_l^Z = -\frac{\|ZU\|_{\mathrm{HS}}^2}{b_l} + \frac{\|ZU\|_{\mathrm{HS}}^2}{b_{l+1}} = \delta \frac{\|ZU\|_{\mathrm{HS}}^2}{b_l b_{l+1}},$$

so taking into account the fact that $\Delta_l \geqslant \Delta_l^Z \geqslant 0$, it remains to prove the following :

$$\frac{\|ZU\|_{\mathrm{HS}}^2}{b_{l+1}^2} \leqslant \delta^2 \frac{\|ZU\|_{\mathrm{HS}}^4}{\|U\|_2^2 b_l^2 b_{l+1}^2} + \frac{\|ZU\|_{\mathrm{HS}}^2}{b_l b_{l+1}}. \tag{1.9}$$

By Hypothesis (1.7), the inequality (1.9) follows by

$$\frac{\|ZU\|_{\mathrm{HS}}^2}{b_{l+1}^2} \leqslant \delta \frac{\|ZU\|_{\mathrm{HS}}^2}{b_l b_{l+1}^2} + \frac{\|ZU\|_{\mathrm{HS}}^2}{b_l b_{l+1}}, \tag{1.10}$$

which is trivially true since $b_{l+1} = b_l - \delta$. □

We are now able to complete the proof of Theorem 1.4. To this end, we must verify that conditions (1.6) and (1.7) hold at each step. At the beginning we have $A_0 = 0$ and $Z = Id$, so we must choose a barrier b_0 such that :

$$-\frac{\|U\|_{\mathrm{HS}}^2}{b_0} \leqslant -\|D\|_{\mathrm{HS}}^2 - \frac{\|U\|^2}{\delta} \tag{1.11}$$

and

$$b_0 \leqslant \delta \frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2}. \tag{1.12}$$

We choose

$$b_0 := \varepsilon \frac{\|U\|_{\mathrm{HS}}^2}{\|D\|_{\mathrm{HS}}^2} \quad \text{and} \quad \delta := \frac{\varepsilon}{1-\varepsilon} \frac{\|U\|^2}{\|D\|_{\mathrm{HS}}^2},$$

and we note that (1.11) and (1.12) are verified. Also, at each step (1.6) holds because $\phi(A_{l+1}, b_{l+1}) \leqslant \phi(A_l, b_l)$. Since $\|ZU\|_{\mathrm{HS}}^2$ decreases at each step by at most $\|U\|^2$, the right-hand side of (1.7) decreases by at most δ, and therefore (1.7) holds once we replace b_l by $b_l - \delta$.

Finally note that, after $k = (1-\varepsilon)^2 \frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2}$ steps, the barrier will be

$$b_k = b_0 - k\delta = \varepsilon^2 \frac{\|U\|_{\mathrm{HS}}^2}{\|D\|_{\mathrm{HS}}^2}.$$

This completes the proof.

Remark 1.7. *The proof can be a little bit simplified using an idea of Casazza [21] which we will discuss here. However, with this approach we lose a constant factor in the size of the set σ which turns out to be crucial in applications.*
At the beginning, the method is the same and consists of finding a vector v satisfying (1.4).

This is guaranteed by taking the sum and proving (1.5), namely :

$$\operatorname{Tr}(U^t(A_l - b_{l+1}I)^{-2}U) \leqslant \frac{\phi(A_l, b_l) - \phi(A_l, b_{l+1})}{\|U\|_2^2}\left(-\|D\|_{\mathrm{HS}}^2 - \phi(A_l, b_{l+1})\right).$$

While we followed Spielman-Srivastava's idea in Lemma 1.6 to prove (1.5), we could have used a remark of Casazza [21] and prove the following :

Lemma 1.8. *Suppose that A_l has l nonzero eigenvalues all greater than b_l. If $\delta \leqslant \frac{b_l}{2}$ (which means $\delta \leqslant b_{l+1} = b_l - \delta$) and*

$$\phi(A_l, b_l) \leqslant -\|D\|_{\mathrm{HS}}^2 - \frac{2\|U\|^2}{\delta} \tag{1.13}$$

then (1.5) holds and therefore the conclusion of Lemma 1.6 follows.

Proof. The main trick is that we have

$$(A_l - b_l I)^{-1} - (A_l - b_{l+1}I)^{-1} \succeq \frac{\delta}{2}(A_l - b_{l+1}I)^{-2} \tag{1.14}$$

This was previously proven only on the image of A_l without the constant $\frac{1}{2}$. To see this, we may compare the corresponding eigenvalues.

– On the Kernel of A_l.

$$\begin{aligned}\frac{1}{-b_l} - \frac{1}{-b_{l+1}} &= \frac{\delta}{b_l \cdot b_{l+1}} \\ &= \frac{\delta}{b_{l+1}(b_{l+1}+\delta)} \\ &\geqslant \frac{\delta}{2b_{l+1}^2} \qquad \text{since } \delta \leqslant b_{l+1}\end{aligned}$$

– On the Image of A_l, denoting λ_i the nonzero eigenvalues of A_l, we have :

$$\begin{aligned}\frac{1}{\lambda_i - b_l} - \frac{1}{\lambda_i - b_{l+1}} &= \frac{\delta}{(\lambda_i - b_l)(\lambda_i - b_{l+1})} \\ &\geqslant \frac{\delta}{(\lambda_i - b_{l+1})^2} \\ &\geqslant \frac{\delta}{2(\lambda_i - b_{l+1})^2}\end{aligned}$$

This proves (1.14). Multiplying by U^t and U from the two sides in (1.14) and taking the trace we get

$$\phi(A_l, b_l) - \phi(A_l, b_{l+1}) \geqslant \frac{\delta}{2}\mathrm{Tr}\left(U^t(A_l - b_{l+1}I)^{-2}U\right)$$

This means that

$$\mathrm{Tr}\left(U^t(A_l - b_{l+1}I)^{-2}U\right) \leqslant (\phi(A_l, b_l) - \phi(A_l, b_{l+1})) \cdot \frac{2}{\delta} \tag{1.15}$$

Note that (1.13) can be written

$$\frac{2}{\delta} \leqslant \frac{-\|D\|_{\mathrm{HS}}^2 - \phi(A_l, b_l)}{\|U\|^2} \leqslant \frac{-\|D\|_{\mathrm{HS}}^2 - \phi(A_l, b_{l+1})}{\|U\|^2} \tag{1.16}$$

Replacing (1.16) in (1.15), we get (1.5) and finish the proof of the lemma. □

Now it remains to take parameters with respect to this argument. For that, we must choose b_0 and δ satisfying (1.13). Denote k the number of steps to be done. Since at each step, we must have $\delta < b_{l+1}$ then the final condition would be

$$\delta < b_k = b_0 - k\delta \quad i.e \quad \delta < \frac{b_0}{k+1}$$

At the beginning, (1.13) gives the following :

$$-\frac{\|U\|_{\mathrm{HS}}^2}{b_0} \leqslant -\|D\|_{\mathrm{HS}} - \frac{2\|U\|^2}{\delta} \tag{1.17}$$

Choosing $\delta = (1-\varepsilon)\frac{b_0}{k}$ (with $\varepsilon < \frac{1}{2}$) and taking equality in (1.17) we get

$$-\frac{\|U\|_{\mathrm{HS}}^2}{b_0} = -\|D\|_{\mathrm{HS}} - \frac{2k \cdot \|U\|^2}{(1-\varepsilon)b_0}$$

After rearrangement, we have

$$b_0 = \frac{\|U\|_{\mathrm{HS}}^2 - \frac{2k\cdot\|U\|^2}{1-\varepsilon}}{\|D\|_{\mathrm{HS}}^2}$$

In order to get a non trivial information, we need to keep $b_0 > 0$ which means that

$$\|U\|_{\mathrm{HS}}^2 > \frac{2k \cdot \|U\|^2}{1-\varepsilon}$$

Therefore one is led to take

$$k = \frac{(1-\varepsilon)^2}{2} \cdot \frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2}$$

Finally, after k steps, the final barrier will be equal to $b_0 - k\delta = \varepsilon b_0 = \varepsilon^2 \frac{\|U\|_{\mathrm{HS}}^2}{\|D\|_{\mathrm{HS}}^2}$.

The inconvenient of this method, is that we are not allowed to extract more than half of the stable rank. In applications, as we will see in chapter 3, extracting all the quantity is needed and it turns out that a proportion of the stable rank is not even sufficient.

Remark 1.9. *Let us mention that our proof of Theorem 1.4 produces a deterministic algorithm :*

Input: An $n \times m$ matrix U, an $m \times m$ diagonal matrix $D = diag(\alpha_1, ..., \alpha_m)$ and $\varepsilon \in (0, 1)$.

$A_0 = 0 \in \mathcal{M}_{n\times n}$;
$\sigma = \emptyset$;
$b_0 = \varepsilon \frac{\|U\|_{\mathrm{HS}}^2}{\|D\|_{\mathrm{HS}}^2}$;
$\delta = \frac{\varepsilon}{1-\varepsilon} \frac{\|U\|_{\mathrm{HS}}^2}{\|D\|_{\mathrm{HS}}^2}$;
$b_1 = b_0 - \delta$;
for $l = 0$ **to** $[(1-\varepsilon)^2 \mathrm{srank}(U)]$ **do**
  for $j \in I_D$ *and* $j \notin \sigma$ **do**
    if *(1.4) is satisfied* **then**
      $A_{l+1} = A_l + \left(\frac{Ue_j}{\alpha_j}\right) \cdot \left(\frac{Ue_j}{\alpha_j}\right)^t$;
      $\sigma = \sigma \cup \{j\}$;
      $b_{l+2} = b_{l+1} - \delta$;
      break;
    end
  end
end
return σ

Algorithm 1: Algorithm of Theorem 1.4.

1.3 Coordinate projections

Given an $n \times m$ matrix U and an integer $k \leqslant m$, our aim is to find a coordinate projection of U of rank k which gives the best minimal operator norm among all coordinate projections. First results were obtained by Lunin [51] and Kashin [44] and a complete answer to this question was given by Kashin-Tzafriri [43] who proved the following :

Theorem 1.10 (Kashin-Tzafriri)**.** *Let U be an $n \times m$ matrix. Fix λ with $1/m \leqslant \lambda \leqslant \frac{1}{4}$. Then, there exists a subset ν of $\{1, \dots, m\}$ of cardinality $|\nu| \geqslant \lambda m$ such that*

$$\|U_\nu\| \leqslant c\left(\sqrt{\lambda}\|U\|_2 + \frac{\|U\|_{\mathrm{HS}}}{\sqrt{m}}\right),$$

where $U_\nu = UP_\nu^t$ and P_ν denotes the coordinate projection onto $\mathbb{R}^\nu$.

The conclusion of the Theorem states that for a fixed $\lambda \leqslant \frac{1}{4}$ we have

$$\min_{\substack{\sigma \subset \{1,\dots,m\} \\ |\sigma| = \lambda m}} \|U_\sigma\| \leqslant c\left(\sqrt{\lambda}\|U\| + \frac{\|U\|_{\mathrm{HS}}}{\sqrt{m}}\right), \tag{1.18}$$

and this estimate is optimal in the sense that the dependence on the parameters in the right hand side cannot be improved.

Kashin-Tzafriri's proof (see [90]) uses the selectors with some other probabilistic arguments and then the Grothendieck's factorization Theorem (see [27] and [61]). In [87], Tropp gave a randomized algorithm to realize Grothendieck's factorization theorem and therefore he was able to give a randomized algorithm to find the subset σ promised in Theorem 1.10.

Our aim here is to give a deterministic algorithm to find the subset σ. Our method uses tools from the work of Batson-Spielman-Srivastava [13] and allows us to improve Kashin-Tzafriri's result by extending the size of the coordinate projection and getting better constants in the result.

Theorem 1.11. *Let U be an $n \times m$ matrix and let $1/m \leqslant \lambda \leqslant \eta < 1$. Then, there exists $\sigma \subset \{1, \dots, m\}$ with $|\sigma| = k \geqslant \lambda m$ such that*

$$\|U_\sigma\| \leqslant \frac{1}{\sqrt{1-\lambda}}\left(\sqrt{\lambda+\eta}\|U\| + \sqrt{1+\frac{\lambda}{\eta}}\frac{\|U\|_{\mathrm{HS}}}{\sqrt{m}}\right),$$

In particular,

$$\|U_\sigma\| \leqslant \frac{\sqrt{2}}{\sqrt{1-\lambda}}\left(\sqrt{\lambda}\|U\| + \frac{\|U\|_{\mathrm{HS}}}{\sqrt{m}}\right),$$

where U_σ denotes the selection of the columns of U with indices in σ.

Proof. We denote by $(e_j)_{j \leqslant m}$ the canonical basis of $\mathbb{R}^m$. Since

$$U_\sigma \cdot U_\sigma^t = \sum_{j \leqslant \sigma} (Ue_j) \cdot (Ue_j)^t,$$

our problem reduces to the question of estimating the largest eigenvalue of this sum of rank one matrices. We will follow the same procedure as in the proof of the restricted invertibility theorem : at each step, we would like to add a column of the original matrix and then study the evolution of the largest eigenvalue. However, it will be convenient for us to add suitable multiples of the columns of U in order to construct the l-th matrix; for each l we will choose a subset σ_l of cardinality $|\sigma_l| = l$ and consider the matrix $A_l = \sum_{j\in\sigma_l} s_j\,(Ue_j)\cdot(Ue_j)^t$ where $(s_j)_{j\in\sigma}$ will be positive numbers which will be suitably chosen. At the step l, the barrier will be denoted by u_l, namely the eigenvalues of A_l will be all smaller than u_l. The corresponding potential is $\psi(A_l, u_l) := \operatorname{Tr}\left(U^t(u_lI - A_l)^{-1}U\right)$. We set $A_0 = 0$, while u_0 will be determined later.

As we did before, at each step the value of the potential $\psi(A_l, u_l)$ will decrease so that we can continue the iteration, while the value of the barrier will increase by a constant δ, i.e. $u_{l+1} = u_l + \delta$. We will use a lemma which appears as Lemma 3.4 in [78]. We state it here in the notation introduced above.

Lemma 1.12. *Assume that $\lambda_{\max}(A_l) \leqslant u_l$. Let v be a vector in $\mathbb{R}^n$ satisfying*

$$F_l(v) := \frac{v^t(u_{l+1}I - A_l)^{-2}v}{\psi(A_l,u_l) - \psi(A_l,u_{l+1})}\|U\|^2 + v^t(u_{l+1}I - A_l)^{-1}v \leqslant \frac{1}{s}.$$

Then, if we define $A_{l+1} = A_l + svv^t$ we have

$$\lambda_{\max}(A_{l+1}) \leqslant u_{l+1} \quad \text{and} \quad \psi(A_{l+1},u_{l+1}) \leqslant \psi(A_l,u_l).$$

Proof. Using Sherman-Morrison formula we have :

$$\begin{aligned}
\psi(A_{l+1},u_{l+1}) &= \operatorname{Tr}\left(U^t\left(u_{l+1}I - A_l - svv^t\right)U\right)\\
&= \operatorname{Tr}\left(U^t\left(u_{l+1}I - A_l\right)U\right) + \frac{sv^t(u_{l+1}I - A_l)^{-1}UU^t(u_{l+1}I - A_l)^{-1}v}{1 - sv^t(u_{l+1}I - A_l)^{-1}v}\\
&\leqslant \psi(A_l,u_l) - (\psi(A_l,u_l) - \psi(A_l,u_{l+1})) + \frac{v^t(u_{l+1}I - A_l)^{-2}v}{\frac{1}{s} - v^t(u_{l+1}I - A_l)^{-1}v}\|U\|^2
\end{aligned}$$

Since $v^t(u_{l+1}I - A_l)^{-1}v < F_l(v)$ and $F_l(v) \leqslant \frac{1}{s}$ we deduce that the quantity above is finite. This implies that $\lambda_{\max}(A_{l+1}) < u_{l+1}$, since otherwise one would find $s' < s$ such that $\lambda_{\max}(A_l + s'vv^t) = u_{l+1}$ and therefore $\psi(A_l + s'vv^t, u_{l+1})$ would blow up which contradicts the fact that it is finite.

On the other hand, rearranging the inequality above using the fact that $F_l(v) \leqslant \frac{1}{s}$ we get $\psi(A_{l+1}, u_{l+1}) \leqslant \psi(A_l, u_l)$. □

We write α for the initial potential, i.e. $\alpha = \frac{\|U\|_{\mathrm{HS}}^2}{u_0}$. Suppose that $A_l = \sum_{j\in\sigma_l} s_j\,(Ue_j)\cdot(Ue_j)^t$ is constructed so that $\psi(A_l, u_l) \leqslant \psi(A_{l-1}, u_{l-1}) \leqslant \alpha$ and $\lambda_{\max}(A_l) \leqslant u_l$. We will now use Lemma 1.12 in order to construct A_{l+1}. To this end, we must find a vector Ue_j not chosen before and a scalar s_{l+1} so that $F_l(Ue_j) \leqslant \frac{1}{s_{l+1}}$, and then use the lemma. Since $(u_lI - A_l)^{-1}$ and $(u_{l+1}I - A_l)^{-1}$ are positive semidefinite, one can easily check that

$$(u_lI - A_l)^{-1} - (u_{l+1}I - A_l)^{-1} \succeq \delta(u_{l+1}I - A_l)^{-2}.$$

Therefore,

$$\mathrm{Tr}\left(U^t(u_{l+1}I - A_l)^{-2}U\right) \leqslant \frac{1}{\delta}\left(\psi(A_l, u_l) - \psi(A_l, u_{l+1})\right).$$

It follows that

$$\begin{aligned}\sum_{j\notin\sigma_l} F_l(Ue_j) \leqslant \sum_{j\leqslant m} F_l(Ue_j) &= \frac{Tr\left(U^t(u_{l+1}I - A_l)^{-2}U\right)}{\psi(A_l, u_l) - \psi(A_l, u_{l+1})}\|U\|^2 + \psi(A_l, u_{l+1}) \\ &\leqslant \frac{\|U\|^2}{\delta} + \alpha,\end{aligned}$$

and therefore one can find $i \notin \sigma_l$ such that

$$F_l(Ue_i) \leqslant \frac{1}{|\sigma_l^c|}\left(\frac{\|U\|^2}{\delta} + \alpha\right) \leqslant \frac{1}{|\sigma_k^c|}\left(\frac{\|U\|^2}{\delta} + \alpha\right),$$

where k is the maximum number of steps (which is in our case λm).
We are going to choose all s_j equal to $s := \frac{(1-\lambda)m}{\alpha+\frac{\|U\|^2}{\delta}}$. By the previous lemma, it is sufficient to take $A_{l+1} = A_l + s\,(Ue_i)\cdot(Ue_i)^t$. After $k = \lambda m$ steps, we get $\sigma = \sigma_k$ such that

$$\begin{aligned}\lambda_{\max}\left(\sum_{j\in\sigma_k}(Ue_j)\cdot(Ue_j)^t\right) &\leqslant \frac{1}{s}(u_0 + k\delta) = \frac{\alpha + \frac{\|U\|^2}{\delta}}{(1-\lambda)m}(u_0 + k\delta) \\ &= \frac{1}{1-\lambda}\left[\frac{\|U\|_{\mathrm{HS}}^2}{m} + \lambda\|U\|^2 + \lambda\|U\|_{\mathrm{HS}}^2\frac{\delta}{u_0} + \frac{\|U\|^2}{m}\frac{u_0}{\delta}\right]\end{aligned}$$

The result follows by taking $u_0 = \eta m\delta$. The second part of the theorem follows by taking $\lambda = \eta$. □

Remark 1.13. *The proof of Theorem 1.11 produces a deterministic algorithm :*

Input: An $n \times m$ matrix U and $0 < \eta \leqslant \lambda < 1$.

$A_0 = 0 \in \mathcal{M}_{n\times n}$;
$\sigma = \emptyset$;
$u_0 = \eta m$;
$u_1 = u_0 + 1$;
$s = \frac{(1-\lambda)m}{\|U\|^2 + \frac{\|U\|_{\mathrm{HS}}^2}{u_0}}$;
for $l = 0$ **to** $[\lambda m]$ **do**
  for $j = 1$ **to** m *and* $j \notin \sigma$ **do**
    if $F_l(Ue_j) \leqslant s^{-1}$ **then**
      $A_{l+1} = A_l + s\,(Ue_j)\cdot(Ue_j)^t$;
      $\sigma = \sigma \cup \{j\}$;
      $u_{l+2} = u_{l+1} + 1$;
      break;
    end
  end
end
return σ

Algorithm 2: Algorithm of Theorem 1.11.

1.4 Extracting square submatrix with small norm : first attempt

Till now, we always managed to extract columns of the matrix. One can ask about extracting a "large" square submatrix with small norm. Bourgain-Tzafriri [19] answered this problem by proving the following :

Theorem 1.14. *There is a universal constant $c > 0$ such that for every $\varepsilon > 0$ and $n \in \mathbb{N}$, if an operator $T : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $\langle Te_i, e_i\rangle = 0$ for all $i \in \{1, \dots, n\}$ then there exists a subset $\sigma \subseteq \{1, \dots, n\}$ satisfying $|\sigma| \geqslant c\varepsilon^2 n$ and $\|P_\sigma T P_\sigma^t\| \leqslant \varepsilon\|T\|$.*

In [14], it has been proven that the quadratic dependance on ε in the previous statement is optimal. Let us note that Theorem 1.14 implies the restricted invertibility as stated by Bourgain-Tzafriri in Theorem 1.1 with a worst dependance on the norm of T.
To see this, take T a linear operator on l_2^n such that $\|Te_j\|_2 = 1$ for all $j \leqslant n$, where $(e_j)_{j\leqslant n}$ denotes the canonical basis of $\mathbb{R}^n$. Note that $\|T\|^2 \geqslant 1$. Now let $A = T^tT - Id$, then $\langle Ae_i, e_i\rangle = 0$ for all $i \in \{1, \dots, n\}$ and $\|A\| \leqslant 1 + \|T\|^2 \leqslant 2\|T\|^2$. Apply Theorem 1.14 to A to find $\sigma \subseteq$

$\{1,\ldots,n\}$ satisfying $|\sigma|\geqslant c\varepsilon^2 n$ and $\|P_\sigma A P_\sigma^t\|\leqslant 2\varepsilon\|T\|^2$. This means that

$$-2\varepsilon\|T\|^2 Id \preceq P_\sigma A P_\sigma^t \preceq 2\varepsilon\|T\|^2 Id$$

and therefore

$$\left(1-2\varepsilon\|T\|^2\right) Id \preceq P_\sigma T^t T P_\sigma^t \preceq \left(1+2\varepsilon\|T\|^2\right) Id$$

Taking $\varepsilon=\frac{1}{4\|T\|^2}$, we get σ of size $c\frac{n}{\|T\|^4}$ such that $s_{\min}(T_\sigma)\geqslant\frac{1}{2}$, which is the statement of Theorem 1.1.

The proof of Theorem 1.14 relies on some probabilistic arguments and Grothendieck's factorization theorem (see [27] and [61]). In [55], Naor asked if one can give a proof of Theorem 1.14 using tools of the method of Batson-Spielman-Srivastava [13] . In other words, find a deterministic algorithm, based on the method of Batson-Spielman-Srivastava, to extract the square submatrix. In this chapter, our proof gives an affirmative answer to this question in the case of symmetric matrices but with a worse dependence on ε.

Proposition 1.15. *Let T be an $n\times n$ symmetric matrix with zero diagonal. There exists $\sigma\subset\{1,..,n\}$ of cardinality $c\varepsilon^4 n$ such that $\|P_\sigma T P_\sigma^t\|\leqslant\varepsilon\|T\|$.*

Proof. Let $A=T+\|T\|\cdot Id$. Clearly A is positive semi-definite, so we can consider $U=A^{\frac{1}{2}}$. Denote by $(e_i)_{i\leqslant n}$ the canonical basis of $\mathbb{R}^n$.

$$\begin{aligned}
\|U\|_{\mathrm{HS}}^2 &= \sum_{i=1}^n \|Ue_i\|_2^2\\
&= \sum_{i=1}^n \langle Ue_i, Ue_i\rangle\\
&= \sum_{i=1}^n \langle Ae_i, e_i\rangle\\
&= \sum_{i=1}^n \langle Te_i, e_i\rangle + \|T\|\sum_{i=1}^n \langle e_i, e_i\rangle\\
&= \mathrm{Tr}(T) + n\|T\|\\
&= n\|T\|
\end{aligned}$$

In addition we have $\|U\|^2=\|A\|=2\|T\|$.

The main idea is to run successively the algorithms of Theorem 1.11 and Theorem 1.4. Let $\varepsilon>0$ and take $\varepsilon_1=\frac{\varepsilon}{3+\varepsilon}$.

Apply Theorem 1.11 to U with $\lambda = \varepsilon_1^2$ and $\eta = \varepsilon_1$ to find $\nu \subset \{1,..,n\}$ of size λn such that

$$\|UP_\nu^t\|^2 \leqslant \frac{1}{1-\varepsilon_1^2}\left(\left(\varepsilon_1^2+\varepsilon_1\right)\|U\|^2 + (1+\varepsilon_1)\frac{\|U\|_{\mathrm{HS}}^2}{n}\right)$$

Replacing the values of $\|U\|^2$ and $\|U\|_{\mathrm{HS}}^2$ we get

$$\|UP_\nu^t\|^2 \leqslant \frac{1}{1-\varepsilon_1^2}\left(2\left(\varepsilon_1^2+\varepsilon_1\right)\|T\| + (1+\varepsilon_1)\|T\|\right)$$

Simplifying the calculation, the previous inequality implies the following

$$P_\nu U^t U P_\nu^t = \left(UP_\nu^t\right)^t \cdot \left(UP_\nu^t\right) \preceq \frac{1+3\varepsilon_1+2\varepsilon_1^2}{1-\varepsilon_1^2}\|T\|Id$$

Now denote $U_1 = UP_\nu^t$. We have just established that

$$\|U_1\|^2 \leqslant \frac{1+3\varepsilon_1+2\varepsilon_1^2}{1-\varepsilon_1^2}\|T\| = \left(1+\frac{3\varepsilon_1}{1-\varepsilon_1}\right)\|T\|$$

We also have

$$\|U_1\|_{\mathrm{HS}}^2 = |\nu|\cdot\|T\| = \varepsilon_1^2 n\|T\|$$

Now apply Theorem 1.4 with U_1, $D = Id$ and parameter $\left(1-\frac{\varepsilon}{2}\right)$ to find $\sigma \subset \nu$ of size

$$\begin{aligned}
\frac{\varepsilon^2}{4}\frac{\|U_1\|_{\mathrm{HS}}^2}{\|U_1\|^2} &= \frac{1}{4}\varepsilon^2\frac{\varepsilon_1^2}{1+\frac{3\varepsilon_1}{1-\varepsilon_1}}\cdot n \\
&= \frac{\varepsilon^4}{4(3+\varepsilon)^2(1+\varepsilon)}\cdot n \\
&\geqslant \frac{\varepsilon^4 n}{128}
\end{aligned}$$

such that

$$P_\sigma U^t U P_\sigma^t = \left(U_1P_\sigma^t\right)^t \cdot \left(U_1P_\sigma^t\right) \succeq \left(1-\frac{\varepsilon}{2}\right)^2\frac{\|U_1\|_{\mathrm{HS}}^2}{\|D\|_{\mathrm{HS}}^2}Id = \left(1-\frac{\varepsilon}{2}\right)^2\|T\|Id$$

Taking in account that $U^tU = A$, we have shown that

$$\left(1-\frac{\varepsilon}{2}\right)^2\|T\|Id \preceq P_\sigma A P_\sigma^t \preceq \left(1+\frac{3\varepsilon_1}{1-\varepsilon_1}\right)\|T\|Id$$

which gives us

$$\left(\frac{\varepsilon^2}{4} - \varepsilon\right)\|T\|Id \preceq P_\sigma T P_\sigma^t \preceq \frac{3\varepsilon_1}{1-\varepsilon_1}\|T\|Id$$

and finally that $\|P_\sigma T P_\sigma^t\| \leqslant \max\left(\frac{3\varepsilon_1}{1-\varepsilon_1}, \varepsilon - \frac{\varepsilon^2}{4}\right)\|T\| = \max\left(\varepsilon, \varepsilon - \frac{\varepsilon^2}{4}\right)\|T\| = \varepsilon\|T\|$.

□

Chapitre 2

Column subset selection

2.1 Introduction

Let U be an $n \times m$ matrix. As we have seen in the previous chapter, the stable rank of U is given by $srank(U) = \frac{\|U\|_{\text{HS}}^2}{\|U\|^2}$, where $\|U\|_{\text{HS}}^2 = \text{Tr}(UU^t)$ denotes the Hilbert-Schmidt norm of U and $\|U\|$ the operator norm of U seen as an operator from l_2^m to l_2^n. Denoting $\widetilde{U}$ the matrix whose columns are obtained by normalizing those of U, our aim is to extract almost $srank(U)$ linearly independent columns of $\widetilde{U}$ and estimate the smallest and the largest singular value of the restricted matrix. This problem is closely related to the restricted invertibility where only an estimate on the smallest singular value is needed. In the previous chapter, we stated in Theorem 1.2 a part of Vershynin's result proved in [90]. Precisely, Vershynin proved the following :

Theorem 2.1 (Vershynin)**.** *Let $Id = \sum_{j\leqslant m} x_j x_j^t$ and let T be a linear operator on ℓ_2^n. For any $\varepsilon \in (0,1)$, one can find $\sigma \subset \{1, \dots, m\}$ with*

$$|\sigma| \geqslant \lfloor (1-\varepsilon)\text{srank}(T) \rfloor = \left\lfloor (1-\varepsilon)\frac{\|T\|_{\text{HS}}^2}{\|T\|^2} \right\rfloor$$

such that for all scalars $(a_j)_{j\in\sigma}$

$$c_1(\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}} \leqslant \left\| \sum_{j\in\sigma} a_j \frac{Tx_j}{\|Tx_j\|_2} \right\|_2 \leqslant c_2(\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}.$$

The normalization on the vectors $(Tx_j)_{j\in\sigma}$ is crucial for some applications and the dependence on ε plays an important role. Vershynin's estimate gives $c(\varepsilon) = \frac{c_1(\varepsilon)}{c_2(\varepsilon)} \approx \varepsilon^{c\log(\varepsilon)}$. Our aim here, is to improve Vershynin's result obtaining simultaneously a restricted invertibility principle and an estimate on the norm of the restricted matrix. Our proof uses tools of the method of Batson-Spielman-Srivastava [13].

The main result of this chapter is the following :

Theorem 2.2. *Let U be an $n \times m$ matrix and denote by $\widetilde{U}$ the matrix whose columns are the columns of U normalized. For any $\varepsilon \in (0,1)$, there exists $\sigma \subset \{1, ..., m\}$ of size*

$$|\sigma| \geqslant \left\lfloor (1-\varepsilon)^2 \mathrm{srank}(U) \right\rfloor = \left\lfloor (1-\varepsilon)^2 \frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2} \right\rfloor$$

such that

$$\frac{\varepsilon}{2-\varepsilon} \leqslant s_{\min}\left(\widetilde{U}_\sigma\right) \leqslant s_{\max}\left(\widetilde{U}_\sigma\right) \leqslant \frac{2-\varepsilon}{\varepsilon}$$

In other terms, for all scalars $(a_j)_{j\in\sigma}$

$$\frac{\varepsilon}{2-\varepsilon}\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}} \leqslant \left\| \sum_{j\in\sigma} a_j \frac{Ue_j}{\|Ue_j\|_2} \right\|_2 \leqslant \frac{2-\varepsilon}{\varepsilon}\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}.$$

Note that the lower bound problem is the restricted invertibility problem treated in Theorem 1.4 while the upper bound problem is related to the Kashin-Tzafriri column selection theorem treated in Theorem 1.11. Our idea is to merge together the two algorithms obtained in the previous chapter in order to get the two conclusions simultaneously. The heart of these methods is the study of the evolution of the eigenvalues of a matrix when perturbated by a rank one matrix.

In the regime where ε is close to one, the previous result yields the following :

Corollary 2.3. *Let U be an $n \times m$ matrix and denote by $\widetilde{U}$ the matrix whose columns are the columns of U normalized. For any $\varepsilon \in (0,1)$, there exists $\sigma \subset \{1, ..., m\}$ of size*

$$|\sigma| \geqslant \left\lfloor \frac{\varepsilon^2}{9} . \frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2} \right\rfloor$$

such that

$$1-\varepsilon \leqslant s_{\min}\left(\widetilde{U}_\sigma\right) \leqslant s_{\max}\left(\widetilde{U}_\sigma\right) \leqslant 1+\varepsilon$$

In other terms, for all scalars $(a_j)_{j\in\sigma}$

$$(1-\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}} \leqslant \left\|\sum_{j\in\sigma} a_j \frac{Ue_j}{\|Ue_j\|_2}\right\|_2 \leqslant (1+\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}.$$

This result is also related to the problem of column paving that is partitioning the columns into sets such that each of the corresponding restrictions has "good" bounds on the singular values, in particular such that the singular values are close to one. We will show how our theorem allows us to recover a result of Tropp [87] (and of Bourgain-Tzafriri [19]) dealing with this problem, using our deterministic method instead of the probabilistic methods used previously.

In a survey [55] on Batson-Spielman-Srivastava's sparsification theorem [13], Naor asked about giving a proof of another theorem of Bourgain-Tzafriri [19], which is stronger than the restricted invertibility, using tools from Batson-Spielman-Srivastava's method. The theorem in question is the following :

Theorem 2.4 (Bourgain-Tzafriri)**.** *There is a universal constant* $c>0$ *such that for every* $\varepsilon\in(0,1)$ *and* $n\in\mathbb{N}$ *if an operator* $T:\mathbb{R}^n\longrightarrow\mathbb{R}^n$ *satisfies* $\langle Te_i,e_i\rangle=0$ *for all* $i\in\{1,\dots,n\}$ *then there exists a subset* $\sigma\subseteq\{1,\dots,n\}$ *satisfying* $|\sigma|\geqslant c\varepsilon^2 n$ *and* $\|P_\sigma T P_\sigma^t\|\leqslant\varepsilon\|T\|$.

We have already tried to solve this question in the previous chapter but our first attempt gave a wrong dependance on ε. Using the main result here, we will be able to give a deterministic algorithm to solve this problem for symmetric matrices. This result being naturally related to the Kadison-Singer conjecture ([41], see also [23]), we also give a deterministic algorithm to recover the best know result on this conjecture for general symmetric matrices.

2.2 Proof of Theorem 2.2

Note $k=|\sigma|=(1-\varepsilon)^2\frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2}$ and

$$A_k=\sum_{j\in\sigma} s_j\left(\widetilde{U}e_j\right)\cdot\left(\widetilde{U}e_j\right)^t,$$

where s_j are positive numbers which will be determined later. Since our aim is to find σ such that the smallest singular value of $\widetilde{U}_\sigma$ is bounded away from zero and its largest one is upper bounded, it is equivalent to try to construct the matrix A_k such that A_k has k eigenvalues bounded away from zero and bounded from above and to estimate the weights s_j. Our construction will be done step by step starting from $A_0 = 0$. So at the beginning, all the eigenvalues of A_0 are zero. At the first step we will try to find a vector v among the columns of $\widetilde{U}$ and a weight s such that $A_1 = A_0 + svv^t$ has one nonzero eigenvalue which have a lower and upper bound. Of course the first step is trivial, since for whatever column we choose, the matrix A_1 will have one eigenvalue equal to s. At the second step, we will try to find a vector v among the columns of $\widetilde{U}$ and a weight s such that $A_2 = A_1 + svv^t$ has two nonzero eigenvalues for which we can update the lower and upper bound found in the first step. We will continue this procedure until we construct the matrix A_k.

For a symmetric matrix A such that $b < \lambda_{\min}(A) \leqslant \lambda_{\max}(A) < u$, we define :

$$\phi(A,b) = \mathrm{Tr}\left(U^t\,(A - b.Id)^{-1}\,U\right) \quad \text{and} \quad \psi(A,u) = \mathrm{Tr}\left(U^t\,(u.Id - A)^{-1}\,U\right)$$

For $l \leqslant k$, we denote by b_l the lower bound of the l nonzero eigenvalues of A_l and by u_l the upper bound i.e $A_l \prec u_l.Id$ and A_l has l eigenvalues $> b_l$. We also note

$$\phi = \phi(A_0, b_0) = -\frac{\|U\|_{\mathrm{HS}}^2}{b_0} \quad \text{and} \quad \psi = \psi(A_0, u_0) = \frac{\|U\|_{\mathrm{HS}}^2}{u_0},$$

where b_0 and u_0 will be determined later.

As we said before, we want to control the evolution of the eigenvalues, so we will make sure to choose a "good" vector so that our bounds b_l and u_l do not move too far. Precisely, we will fix this amount of change and denote it by δ for the lower bound and Δ for the upper bound i.e at the next step the lower bound will be $b_{l+1} = b_l - \delta$ and the upper one will be $u_{l+1} = u_l + \Delta$. We will choose these two quantities as follows :

$$\delta = (1-\varepsilon)\frac{b_0}{k} = \frac{b_0\|U\|^2}{(1-\varepsilon)\|U\|_{\mathrm{HS}}^2} \quad \text{and} \quad \Delta = (1-\varepsilon)\frac{u_0}{k} = \frac{u_0\|U\|^2}{(1-\varepsilon)\|U\|_{\mathrm{HS}}^2}$$

Our choice of δ is motivated by the fact that after k steps we want the updated lower bound to remain positive but not too small due to some obstructions in the proof. In this case the final lower bound will be

$$b_k = b_{k-1} - \delta = ... = b_0 - k\delta = \varepsilon b_0$$

The choice of Δ is motivated by the fact that we don't want the upper bound to move too far from the initial one and as for δ, we have some obstruction on taking its value too small. The final upper bound will be

$$u_k = u_{k-1} + \Delta = ... = u_0 + k\Delta = (2-\varepsilon)u_0$$

Definition 2.5. *We will say that a positive semidefinite matrix A satisfies the l-requirement if the following properties are verified :*

- $A \prec u_l.Id$.
- A *has* l *eigenvalues* $> b_l$.
- $\phi(A, b_l) \leqslant \phi$.
- $\psi(A, u_l) \leqslant \psi$.

In order to construct A_{l+1} which has $l+1$ nonzero eigenvalues larger than b_{l+1} and such that $\phi(A_{l+1}, b_{l+1}) \leqslant \phi(A_l, b_l)$, we may look at the algorithm used for the restricted invertibility problem and more precisely, at the condition needed on the vector v to be chosen. This is basically condition (1.3) appearing in the proof of Theorem 1.4.

Lemma 2.6. *If A_l has l nonzero eigenvalues greater than b_l and if for some vector v and some positive scalar s we have*

$$G_l(v) := -\frac{v^t\left(A_l - b_{l+1}.Id\right)^{-2} v}{\phi(A_l, b_l) - \phi(A_l, b_{l+1})} \cdot \|U\|^2 - v^t\left(A_l - b_{l+1}.Id\right)^{-1} v \geqslant \frac{1}{s}, \tag{2.1}$$

then $A_{l+1} = A_l + svv^t$ has $l+1$ nonzero eigenvalues all greater than b_{l+1} and $\phi(A_{l+1}, b_{l+1}) \leqslant \phi(A_l, b_l)$.

Now, in order to construct A_{l+1} which has all its eigenvalues smaller than u_{l+1} and such that $\psi(A_{l+1}, u_{l+1}) \leqslant \psi(A_l, u_l)$, we may look at the algorithm used for the Kashin-Tzafriri column selection theorem. This is basically Lemma 1.12 appearing in the proof of Theorem 1.11.

Lemma 2.7. *If $A_l \prec u_l.Id$ and if for some vector v and some positive scalar s we have*

$$F_l(v) := \frac{v^t\left(u_{l+1}.Id - A_l\right)^{-2} v}{\psi(A_l, u_l) - \psi(A_l, u_{l+1})} \cdot \|U\|^2 + v^t\left(u_{l+1}.Id - A_l\right)^{-1} v \leqslant \frac{1}{s}. \tag{2.2}$$

Then denoting $A_{l+1} = A_l + svv^t$, we have $A_{l+1} \prec u_{l+1}.Id$ and $\psi(A_{l+1}, u_{l+1}) \leqslant \psi(A_l, u_l)$.

For our problem, we will need to find a vector v satisfying (2.1) and (2.2) simultaneously. For that we need to merge these two conditions in one equation :

Lemma 2.8. *If A_l satisfies the l-requirement and if for some vector v we have*

$$F_l(v) \leqslant G_l(v) \tag{2.3}$$

Then taking any s such that $F_l(v) \leqslant \frac{1}{s} \leqslant G_l(v)$, then $A_{l+1} = A_l + svv^t$ satisfies the $(l+1)$-requirement.

Remark 2.9. *Since A_{l+1} has $l+1$ nonzero eigenvalues while A_l had only l nonzero eigenvalues, then the vector v chosen is linearly independent with the eigenvectors of A_l. Therefore one can see that $\mathrm{Ker}(A_{l+1}) \subset \mathrm{Ker}(A_l)$ and $\mathrm{Dim}\left[\mathrm{Ker(A_{l+1})}\right] = \mathrm{Dim}\left[\mathrm{Ker(A_l)}\right] - 1$.*

Proposition 2.10. *Let A_l satisfying the l-requirement. If b_0 and u_0 satisfy*

$$b_0 \leqslant \frac{\varepsilon u_0}{2-\varepsilon} \tag{2.4}$$

then there exists $i \leqslant m$ and a positive number s_i such that $A_{l+1} = A_l + s_i \left(\tilde{U}e_i\right)\cdot\left(\tilde{U}e_i\right)^t$ satisfies the $(l+1)$-requirement.

Proof. According to Lemma 2.8, it is sufficient to find $i \leqslant m$ such that $F_l\left(\tilde{U}e_i\right) \leqslant G_l\left(\tilde{U}e_i\right)$ and then take s_j such that

$$F_l\left(\tilde{U}e_i\right) \leqslant \frac{1}{s_j} \leqslant G_l\left(\tilde{U}e_i\right) \tag{2.5}$$

Since F_l and G_l are quadratic forms, it is equivalent to find $i \leqslant m$ such that $F_l\left(Ue_i\right) \leqslant G_l\left(Ue_i\right)$. For that, it is sufficient to prove that

$$\sum_{j\leqslant m} F_l\left(Ue_j\right) \leqslant \sum_{j\leqslant m} G_l\left(Ue_j\right) \tag{2.6}$$

Before estimating $\sum_{j\leqslant m} F_l\left(Ue_j\right)$, let us note that

$$\begin{aligned}\psi(A_l,u_l) - \psi(A_l,u_{l+1}) &= \mathrm{Tr}\left[U^t\left(u_l.Id - A_l\right)^{-1}U\right] - \mathrm{Tr}\left[U^t\left(u_{l+1}.Id - A_l\right)^{-1}U\right] \\ &= \Delta\mathrm{Tr}\left[U^t\left(u_l.Id - A_l\right)^{-1}\left(u_{l+1}.Id - A_l\right)^{-1}U\right] \\ &\geqslant \Delta\mathrm{Tr}\left[U^t\left(u_{l+1}.Id - A_l\right)^{-2}U\right]\end{aligned}$$

Replacing this in F_l we get

$$\begin{aligned}
\sum_{j\leqslant m} F_l\left(Ue_j\right) &= \frac{\sum_{j\leqslant m} e_j^t U^t \left(u_{l+1}.Id - A_l\right)^{-2} Ue_j}{\psi(A_l,u_l)-\psi(A_l,u_{l+1})}\cdot \|U\|^2 + \sum_{j\leqslant m} e_j^t U^t \left(u_{l+1}.Id - A_l\right)^{-1} Ue_j \\
&= \frac{\mathrm{Tr}\left[U^t \left(u_{l+1}.Id - A_l\right)^{-2} U\right]}{\psi(A_l,u_l)-\psi(A_l,u_{l+1})}\cdot \|U\|^2 + \mathrm{Tr}\left[U^t \left(u_{l+1}.Id - A_l\right)^{-1} U\right] \\
&\leqslant \frac{\|U\|^2}{\Delta} + \psi
\end{aligned}$$

Now we may estimate $\sum_{j\leqslant m} G_l\left(Ue_j\right)$. We denote by P_l the orthogonal projection onto the image of A_l and Q_l the orthogonal projection onto the kernel of A_l. Note that for any $l\leqslant k$ we have the following fact

$$b_l \leqslant \delta \frac{\|Q_lU\|_{\mathrm{HS}}^2}{\|U\|^2} \tag{2.7}$$

Since $Q_0 = Id$, this fact is true at the beginning by our choice of δ. Taking in account Remark 2.9, at each step $\|Q_lU\|_{\mathrm{HS}}^2$ decreases by at most $\|U\|^2$ so that the right hand side of (2.7) decreases by at most δ. Since at each step we replace b_l by b_{l+1}, (2.7) remains true.
Since $Id = P_l + Q_l$ and P_l, Q_l, A_l commute we can write

$$\begin{aligned}
\mathrm{Tr}\left[U^t \left(A_l - b_{l+1}.Id\right)^{-2} U\right] &= \mathrm{Tr}\left[U^t P_l \left(A_l - b_{l+1}.Id\right)^{-2} P_l U\right] + \mathrm{Tr}\left[U^t Q_l \left(A_l - b_{l+1}.Id\right)^{-2} Q_l U\right] \\
&= \mathrm{Tr}\left[U^t P_l \left(A_l - b_{l+1}.Id\right)^{-2} P_l U\right] + \frac{\|Q_lU\|_{\mathrm{HS}}^2}{b_{l+1}^2}
\end{aligned}$$

Doing the same decomposition for $\phi(A_l,b_l)$ we get

$$\begin{aligned}
\phi(A_l,b_l) &= \mathrm{Tr}\left[U^t P_l \left(A_l - b_l.Id\right)^{-1} P_l U\right] + \mathrm{Tr}\left[U^t Q_l \left(A_l - b_l.Id\right)^{-1} Q_l U\right] \\
&= \mathrm{Tr}\left[U^t P_l \left(A_l - b_l.Id\right)^{-1} P_l U\right] - \frac{\|Q_lU\|_{\mathrm{HS}}^2}{b_l} \\
&:= \phi^P(A_l,b_l) + \phi^Q(A_l,b_l)
\end{aligned}$$

Denote $\Lambda_l = \phi(A_l,b_l) - \phi(A_l,b_{l+1})$ and Λ_l^P, Λ_l^Q the corresponding decompositions onto the image part and the kernel part as above. As we did before, we have $\Lambda_l = \Lambda_l^P + \Lambda_l^Q$ and

$$\begin{aligned}
\Lambda_l^P &= \mathrm{Tr}\left[U^t P_l \left(A_l - b_l.Id\right)^{-1} P_l U\right] - \mathrm{Tr}\left[U^t P_l \left(A_l - b_{l+1}.Id\right)^{-1} P_l U\right] \\
&= \delta \mathrm{Tr}\left[U^t P_l \left(A_l - b_l.Id\right)^{-1} \left(A_l - b_{l+1}.Id\right)^{-1} P_l U\right] \\
&\geqslant \delta \mathrm{Tr}\left[U^t P_l \left(A_l - b_{l+1}.Id\right)^{-2} P_l U\right]
\end{aligned}$$

Using (2.7) we have

$$\Lambda_l^Q = -\frac{\|Q_l U\|_{\mathrm{HS}}^2}{b_l} + \frac{\|Q_l U\|_{\mathrm{HS}}^2}{b_{l+1}} = \frac{\delta\|Q_l U\|_{\mathrm{HS}}^2}{b_l b_{l+1}} \geqslant \frac{\|U\|^2}{b_{l+1}}$$

Looking at the previous informations, we can write

$$\begin{aligned}
\sum_{j\leqslant m} G_l\left(Ue_j\right) &= -\frac{\mathrm{Tr}\left[U^t\left(A_l - b_{l+1}.Id\right)^{-2}U\right]}{\Lambda_l}\cdot\|U\|^2 - \mathrm{Tr}\left[U^t\left(A_l - b_{l+1}.Id\right)^{-1}U\right] \\
&= -\frac{\mathrm{Tr}\left[U^t P_l\left(A_l - b_{l+1}.Id\right)^{-2}P_l U\right] + \frac{\|Q_l U\|_{\mathrm{HS}}^2}{b_{l+1}^2}}{\Lambda_l}\cdot\|U\|^2 - \phi(A_l, b_{l+1}) \\
&\geqslant -\frac{\frac{\Lambda_l^P}{\delta} + \frac{\delta\|Q_l U\|_{\mathrm{HS}}^2}{b_l b_{l+1}}\left[\frac{b_l}{\delta b_{l+1}}\right]}{\Lambda_l}\cdot\|U\|^2 + \Lambda_l - \phi(A_l, b_l) \\
&\geqslant -\frac{\frac{\Lambda_l^P}{\delta} + \Lambda_l^Q\left[\frac{1}{\delta} + \frac{1}{b_{l+1}}\right]}{\Lambda_l}\cdot\|U\|^2 + \Lambda_l^Q - \phi \\
&\geqslant -\frac{\|U\|^2}{\delta} - \frac{\|U\|^2}{b_{l+1}} + \Lambda_l^Q - \phi \\
&\geqslant -\frac{\|U\|^2}{\delta} - \phi
\end{aligned}$$

Until now we have proven that

$$\sum_{j\leqslant m} G_l\left(Ue_j\right) \geqslant -\frac{\|U\|^2}{\delta} - \phi \quad \text{and} \quad \sum_{j\leqslant m} F_l\left(Ue_j\right) \leqslant \frac{\|U\|^2}{\Delta} + \psi$$

So in order to prove (2.6), it will be sufficient to verify

$$\frac{\|U\|^2}{\Delta} + \psi \leqslant -\frac{\|U\|^2}{\delta} - \phi \tag{2.8}$$

Replacing in (2.8) the values of the corresponding parameters as chosen at the beginning, it is sufficient to prove that

$$\frac{(2-\varepsilon)\|U\|_{\mathrm{HS}}^2}{u_0} \leqslant \frac{\varepsilon\|U\|_{\mathrm{HS}}^2}{b_0},$$

which is after rearrangement condition (2.4). □

Keeping in mind that $k = (1-\varepsilon)^2\frac{\|U\|_{\mathrm{HS}}^2}{\|U\|^2}$, we are ready to finish the construction of A_k. We may iterate Proposition 2.10 starting with $A_0 = 0$. Of course, A_0 satisfies the 0-requirement so by the proposition we can find a column vector and a corresponding scalar to form A_1 satisfying the

1-requirement. Once again we use the proposition to construct A_2 satisfying the 2-requirement. We can continue with this procedure as long as the corresponding lower bound b_l is positive (which is the case for b_k). So after k steps we have constructed $A_k = \sum_{j\in\sigma} s_j \left(\tilde{U}e_i\right)\cdot\left(\tilde{U}e_i\right)^t$ satisfying the k-requirement which means that

$$A_k \prec u_k.Id = (2-\varepsilon)u_0 \quad \text{and} \quad A_k \text{ has } k \text{ eigenvalues bigger than } b_k = \varepsilon b_0.$$

Now it remains to estimate the weights s_j chosen. This will be done by a trivial calculation :

Lemma 2.11. *For any $l \leqslant k$ and any unit vector v we have*

$$G_l(v) \leqslant \frac{1}{\varepsilon b_0} \qquad \text{and} \qquad F_l(v) \geqslant \frac{1}{(2-\varepsilon)u_0}$$

Proof. Write $Id = P_l + Q_l$ and notice that $\frac{v^t(A_l - b_{l+1}.Id)^{-2}v}{\phi(A_l,b_l)-\phi(A_l,b_{l+1})}$ and $v^tP_l\left(A_l - b_{l+1}.Id\right)^{-1}P_lv$ are positive, then we have

$$G_l(v) \leqslant -v^t\left(A_l - b_{l+1}.Id\right)^{-1}v \leqslant -v^tQ_l\left(A_l - b_{l+1}.Id\right)^{-1}Q_lv \leqslant \frac{\|Q_lv\|_2^2}{b_{l+1}} \leqslant \frac{\|v\|_2^2}{b_k} \leqslant \frac{1}{\varepsilon b_0}$$

Now since $\frac{v^t(u_{l+1}.Id - A_l)^{-2}v}{\psi(A_l,u_l)-\psi(A_l,u_{l+1})} \geqslant 0$ then

$$F_l(v) \geqslant v^t\left(u_{l+1}.Id - A_l\right)^{-1}v \geqslant v^t\left(u_k.Id\right)^{-1}v \geqslant \frac{1}{(2-\varepsilon)u_0}$$

□

The weights s_j that we have chosen satisfied (2.5) and therefore by the previous lemma

$$\forall i \leqslant k, \quad \varepsilon b_0 \leqslant s_j \leqslant (2-\varepsilon)u_0$$

Back to our problem note that

$$\tilde{U}_\sigma\tilde{U}_\sigma^t = \sum_{j\in\sigma}\left(\tilde{U}e_i\right)\cdot\left(\tilde{U}e_i\right)^t$$

and therefore

$$\frac{1}{(2-\varepsilon)u_0}A_k = \frac{1}{(2-\varepsilon)u_0}\sum_{j\in\sigma}s_j\left(\tilde{U}e_i\right)\cdot\left(\tilde{U}e_i\right)^t \preceq \tilde{U}_\sigma\tilde{U}_\sigma^t \preceq \frac{1}{\varepsilon b_0}\sum_{j\in\sigma}s_j\left(\tilde{U}e_i\right)\cdot\left(\tilde{U}e_i\right)^t = \frac{1}{\varepsilon b_0}A_k$$

Transfering the properties of A_k, we deduce that $\widetilde{U}_\sigma \widetilde{U}_\sigma^t \preceq \frac{(2-\varepsilon)u_0}{\varepsilon b_0} Id$ and $\widetilde{U}_\sigma \widetilde{U}_\sigma^t$ has k eigenvalues greater than $\frac{\varepsilon b_0}{(2-\varepsilon)u_0}$.

This means that

$$\frac{\varepsilon b_0}{(2-\varepsilon)u_0} Id \preceq \widetilde{U}_\sigma^t \widetilde{U}_\sigma \preceq \frac{(2-\varepsilon)u_0}{\varepsilon b_0} Id$$

Taking $b_0 = \frac{\varepsilon u_0}{2-\varepsilon}$ in order to satsify (2.4), we finish the proof of Theorem 2.2.

Remark 2.12. *The proof of Theorem 2.2 produces a deterministic algorithm :*

Input: An $n \times m$ matrix U and $\varepsilon \in (0,1)$.

$A_0 = 0 \in \mathcal{M}_{n\times n}$;
$\sigma = \emptyset$;
$\delta = \frac{\varepsilon}{1-\varepsilon} \frac{\|U\|^2}{\|U\|_{\mathrm{HS}}^2}$;
$\Delta = \frac{2-\varepsilon}{1-\varepsilon} \frac{\|U\|^2}{\|U\|_{\mathrm{HS}}^2}$;
$u_0 = 2-\varepsilon$;
$b_0 = \varepsilon$;
$u_1 = u_0 + \Delta$;
$b_1 = b_0 - \delta$;
for $l = 0$ **to** $[(1-\varepsilon)^2 \mathrm{srank}(U)]$ **do**
  for $j = 1$ **to** m **do**
    if $F_l(\tilde{U}e_j) \leqslant G_l(\tilde{U}e_j)$ **then**
      $s_j = G_l(\tilde{U}e_j)^{-1}$;
      $A_{l+1} = A_l + s_j \left(\tilde{U}e_j\right) \cdot \left(\tilde{U}e_j\right)^t$;
      $\sigma = \sigma \cup \{j\}$;
      $u_{l+2} = u_{l+1} + \Delta$;
      $b_{l+2} = b_{l+1} - \delta$;
      break;
    end
  end
end
return σ

Algorithm 3: Algorithm of Theorem 2.2.

2.3 Application to the local theory of Banach spaces

As for the restricted invertibility principle where one can interpret the result as the invertibility of an operator on a decomposition of the identity, we will write the result in terms of a decomposition of the identity. This will be useful for applications to the local theory of Banach spaces since by John's theorem [40], one can have a decomposition of the identity formed by contact points of the unit ball with its maximal volume ellipsoid.

Proposition 2.13. *Let* $Id = \sum_{i\leqslant m} y_i y_i^t$ *be a decomposition of the identity in* $\mathbb{R}^n$ *and* T *be a linear operator on* l_2^n*. For* $\varepsilon \in (0,1)$*, there exists* $\sigma \subset \{1, ..., m\}$ *such that*

$$|\sigma| \geqslant \left\lfloor (1-\varepsilon)^2 \frac{\|T\|_{\mathrm{HS}}^2}{\|T\|^2} \right\rfloor$$

and for all scalars $(a_j)_{j\in\sigma}$,

$$\frac{\varepsilon}{2-\varepsilon}\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}} \leqslant \left\| \sum_{j\in\sigma} a_j \frac{Ty_j}{\|Ty_j\|_2} \right\|_2 \leqslant \frac{2-\varepsilon}{\varepsilon}\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}.$$

Proof. Let U be the $n \times m$ matrix whose columns are Ty_j. Therefore, we can write

$$UU^t = \sum_{j\leqslant m} (Ty_j)\cdot (Ty_j)^t = TT^t$$

We deduce that $\|U\|_{\mathrm{HS}} = \|T\|_{\mathrm{HS}}$ and $\|U\| = \|T\|$. Applying Theorem 2.2 to U, we find $\sigma \subset \{1, ..., m\}$ such that

$$|\sigma| \geqslant \left\lfloor (1-\varepsilon)^2 \frac{\|T\|_{\mathrm{HS}}^2}{\|T\|^2} \right\rfloor$$

and for all scalars $(a_j)_{j\in\sigma}$,

$$\frac{\varepsilon}{2-\varepsilon}\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}} \leqslant \left\| \sum_{j\in\sigma} a_j \frac{Ue_j}{\|Ue_j\|_2} \right\|_2 \leqslant \frac{2-\varepsilon}{\varepsilon}\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}$$

Noting that $\frac{Ue_j}{\|Ue_j\|_2} = \frac{Ty_j}{\|Ty_j\|_2}$, we finish the proof.

□

This result improves the dependence on ε in comparison with Vershynin's result [90]. While Vershynin proved that $(Ty_j)_{j\in\sigma}$ is $c(\varepsilon)$-equivalent to an orthogonal basis of $\mathbb{R}^\sigma$, the value of $c(\varepsilon)$ was of the order of $\varepsilon^{-c\log(\varepsilon)}$. Here our sequence is $(4\varepsilon^{-2})$-equivalent to an orthogonal basis of $\mathbb{R}^\sigma$.

In the regime where ε is close to one, the previous proposition yields the following :

Corollary 2.14. *Let* $Id = \sum_{i\leqslant m} y_i y_i^t$ *be a decomposition of the identity in* $\mathbb{R}^n$ *and* T *be a linear* *operator on l_2^n. For $\varepsilon \in (0,1)$, there exists $\sigma \subset \{1,...,m\}$ such that*

$$|\sigma| \geqslant \left\lfloor \frac{\varepsilon^2 \|T\|_{\text{HS}}^2}{9\|T\|^2} \right\rfloor$$

and for all scalars $(a_j)_{j\in\sigma}$,

$$(1-\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}} \leqslant \left\| \sum_{j\in\sigma} a_j \frac{Ty_j}{\|Ty_j\|_2} \right\|_2 \leqslant (1+\varepsilon)\left(\sum_{j\in\sigma} a_j^2\right)^{\frac{1}{2}}$$

The two previous results can be written in terms of contact points, let us for instance write the case of $T = Id$. If $X = (\mathbb{R}^n, \|\cdot\|)$ where $\|\cdot\|$ is a norm on $\mathbb{R}^n$ such that B_2^n is the ellipsoid of maximal volume contained in B_X, then by John's theorem [40] one can get an identity decomposition formed by contact points of B_X with B_2^n. Note that the same holds if B_2^n is the ellipsoid of minimal volume containing B_X. Applying Proposition 2.13 we get the following :

Proposition 2.15. *Let $X = (\mathbb{R}^n, \|\cdot\|)$ where $\|\cdot\|$ is a norm on $\mathbb{R}^n$ such that B_2^n is the ellipsoid of maximal volume contained in B_X. For $\varepsilon \in (0,1)$, there exists $x_1,...,x_k$ contact points of B_X with B_2^n such that*

$$k \geqslant \left\lfloor (1-\varepsilon)^2 n \right\rfloor$$

and for all scalars $(a_j)_{j\leqslant k}$,

$$\frac{\varepsilon}{2-\varepsilon}\left(\sum_{j\leqslant k} a_j^2\right)^{\frac{1}{2}} \leqslant \left\| \sum_{j\leqslant k} a_j x_j \right\|_2 \leqslant \frac{2-\varepsilon}{\varepsilon}\left(\sum_{j\leqslant k} a_j^2\right)^{\frac{1}{2}}.$$

In other terms, we can find a system of almost n contact points which is $(4\varepsilon^{-2})$-equivalent to an orthonormal basis. If we are willing to give up on the fact of extracting a large number of contact points, we can have a system of contact points which is $(1+\varepsilon)$-equivalent to an orthonormal basis. For that, we write the previous proposition in the regime where ε is close to 1.

Corollary 2.16. *Let $X = (\mathbb{R}^n, \|\cdot\|)$ where $\|\cdot\|$ is a norm on $\mathbb{R}^n$ such that B_2^n is the ellipsoid of maximal volume contained in B_X. For $\varepsilon \in (0,1)$, there exists $x_1,...,x_k$ contact points of B_X with B_2^n such that*

$$k \geqslant \left\lfloor \frac{\varepsilon^2 n}{9} \right\rfloor$$

and for all scalars $(a_j)_{j\leqslant k}$,

$$(1-\varepsilon)\left(\sum_{j\leqslant k} a_j^2\right)^{\frac{1}{2}} \leqslant \left\|\sum_{j\leqslant k} a_j x_j\right\|_2 \leqslant (1+\varepsilon)\left(\sum_{j\leqslant k} a_j^2\right)^{\frac{1}{2}}$$

Given $X=(\mathbb{R}^n,\|\cdot\|)$, it is proven in [85] that there exist orthonormal vectors $x_1,...,x_k$ with $k=[\frac{\sqrt{n}}{4}]$ such that the three norms $\|\cdot\|_X, \|\cdot\|_{X^*}$ and $\|\cdot\|_2$ differ by a factor 2 on the sequence $(x_j)_{j\leqslant k}$. Corollary 2.16 allows us to find almost orthogonal vectors $x_1,...,x_k$ with $k=[\frac{\varepsilon^2 n}{9}]$ such that the three norms $\|\cdot\|_X, \|\cdot\|_{X^*}$ and $\|\cdot\|_2$ are equal to one on the sequence $(x_j)_{j\leqslant k}$.

Applying proposition 2.15, we also get the following corollary :

Corollary 2.17. *Let* $X=(\mathbb{R}^n,\|\cdot\|)$ *where* $\|\cdot\|$ *is a norm on* $\mathbb{R}^n$ *such that* B_2^n *is the ellipsoid of maximal volume contained in* B_X*. For* $\varepsilon\in(0,1)$*, there exist* $x_1,...,x_k$ *linearly independent contact points of* B_X *with* B_2^n *such that*

$$k \geqslant (\sqrt{2}-1)^4\varepsilon^2 n$$

and for any $i\leqslant k$,

$$\sum_{j\neq i}\langle x_j, x_i\rangle^2 \leqslant \varepsilon$$

Proof. Let $\varepsilon<1$ and denote $\alpha=(\sqrt{2}-1)^2$. Apply proposition 2.15 with $(1-\alpha\varepsilon)$ in order to find a system of linearly independent contact points $(x_j)_{j\leqslant k}$ such that

$$A_k=\sum_{j\leqslant k} x_j x_j^t \preceq \left(\frac{1+\alpha\varepsilon}{1-\alpha\varepsilon}\right)^2 Id$$

For $i\leqslant k$,

$$\langle A_k x_i, x_i\rangle = 1+\sum_{j\neq i}\langle x_j, x_i\rangle^2 \leqslant \left(\frac{1+\alpha\varepsilon}{1-\alpha\varepsilon}\right)^2$$

The conclusion follows by a trivial calculation. □

2.4 Column paving

Extracting a large column submatrix reveals to be useful since the extracted matrix may have better properties. First results in this direction were given by Kashin in [44], and others followed improving or dealing with different properties (see [13], [19], [43], [51],[87]). One can also be interested in partitioning the matrix into disjoint sets of columns such that each block has "good" properties. Obtaining a constant number of blocks (independent of the dimension) turns out to be a difficult problem and many conjectures concerning this were given previously (see [23]).

The previous algorithms for extraction used probabilistic arguments and Grothendieck's factorization theorem (see [27] and [61]). Here we propose a deterministic algorithm to achieve the extraction, we apply our main result iteratively in order to partition the matrix into blocks on each of them we have good estimates on the singular values.

The conditioning number of a matrix U is given by

$$\kappa(U) = \max\left\{\frac{\|Ux\|_2}{\|Uy\|_2};\ \|x\|_2 = \|y\|_2 = 1\right\} = \frac{s_{\max}(U)}{s_{\min}(U)}$$

Clearly, if the matrix is not of full rank then its conditioning number goes to infinity. An interesting problem is to extract a well conditioned submatrix of a given matrix ; by well conditioned submatrix, we mean a submatrix with bounded conditioning number. When the conditioning number is close to one, the matrix is close to a multiple of an isometry. Results of this chapter goes in this direction.

Definition 2.18. *Let U be an $n \times m$ matrix. We will say that U is standardized if all its columns are of norm* 1*.*

Note that when U is standardized we have $\|U\|_{\text{HS}}^2 = m$ and $\|U\| \geqslant 1$. Applying Theorem 2.2 to a standardized matrix, we get the following proposition :

Proposition 2.19. *Let U be an $n \times m$ standardized matrix. For $\varepsilon \in (0,1)$, there exists $\sigma \subset \{1, ..., m\}$ with*

$$|\sigma| \geqslant \left\lfloor \frac{(1-\varepsilon)^2 m}{\|U\|^2} \right\rfloor$$

such that

$$\frac{\varepsilon}{2-\varepsilon} \leqslant s_{\min}(U_\sigma) \leqslant s_{\max}(U_\sigma) \leqslant \frac{2-\varepsilon}{\varepsilon}$$

This means that there exists a "large" submatrix of U whose conditioning number is smaller than $\left(\frac{2-\varepsilon}{\varepsilon}\right)^2$. In the regime where ε is close to one, the previous proposition yields an almost isometric estimation :

Corollary 2.20. *Let U be an $n \times m$ standardized matrix. For $\varepsilon \in (0,1)$, there exists $\sigma \subset \{1, ..., m\}$ with*

$$|\sigma| \geqslant \left\lfloor \frac{\varepsilon^2 m}{9\|U\|^2} \right\rfloor$$

such that

$$1 - \varepsilon \leqslant s_{\min}(U_\sigma) \leqslant s_{\max}(U_\sigma) \leqslant 1 + \varepsilon$$

We may now iterate Proposition 2.19 in order to partition a standardized matrix U into well conditioned blocks.

Proposition 2.21. *Let U be an $n \times m$ standardized matrix. For $\varepsilon \in (0,1)$, there exists a partition of $\{1, ..., m\}$ into p sets $\sigma_1, ..., \sigma_p$ such that*

$$p \leqslant \frac{\|U\|^2 \log(m)}{(1-\varepsilon)^2}$$

and for any $i \leqslant p$,

$$\frac{\varepsilon}{2-\varepsilon} \leqslant s_{\min}(U_{\sigma_i}) \leqslant s_{\max}(U_{\sigma_i}) \leqslant \frac{2-\varepsilon}{\varepsilon}$$

Proof. Apply Proposition 2.19 to U in order to get σ_1 verifying

$$|\sigma_1| \geqslant \frac{(1-\varepsilon)^2}{\|U\|^2} m$$

such that

$$\frac{\varepsilon}{2-\varepsilon} \leqslant s_{\min}(U_{\sigma_1}) \leqslant s_{\max}(U_{\sigma_1}) \leqslant \frac{2-\varepsilon}{\varepsilon}$$

Now note that $U_{\sigma_1^c}$ is an $n \times |\sigma_1^c|$ standardized matrix and $\|U_{\sigma_1^c}\| \leqslant \|U\|$. Apply Proposition 2.19 to $U_{\sigma_1^c}$ in order to get $\sigma_2 \subset \sigma_1^c$ verifying

$$|\sigma_2| = \frac{(1-\varepsilon)^2}{\|U\|^2}|\sigma_1^c| = \frac{(1-\varepsilon)^2}{\|U\|^2}\left[1 - \frac{(1-\varepsilon)^2}{\|U\|^2}\right] m$$

Doing this procedure p times, the number of remaining columns is

$$\left(1 - \frac{(1-\varepsilon)^2}{\|U\|^2}\right)^p m$$

So in order to cover all the columns, we need to take p such that

$$\left(1 - \frac{(1-\varepsilon)^2}{\|U\|^2}\right)^p m < 1$$

By a trivial calculation, it is sufficient to take $\frac{\|U\|^2 \log(m)}{(1-\varepsilon)^2}$ blocks. □

In the regime where ε is close to one, the previous proposition yields a column partition with almost isometric blocks. This recovers a result of Tropp (see Theorem 1.2 in [87]), which follows results of Bourgain-Tzafriri [19], with a deterministic method.

Corollary 2.22. *Let U be an $n \times m$ standardized matrix. For $\varepsilon \in (0,1)$, there exists a partition of $\{1,...,m\}$ into p sets $\sigma_1,...,\sigma_p$ such that*

$$p \leqslant \frac{9\|U\|^2 \log(m)}{\varepsilon^2}$$

and for any $i \leqslant p$,

$$1 - \varepsilon \leqslant s_{\min}\left(U_{\sigma_i}\right) \leqslant s_{\max}\left(U_{\sigma_i}\right) \leqslant 1 + \varepsilon$$

The number of blocks here depends on the dimension. The challenging problem is to partition into a number of blocks which does not depend on the dimension. This would give a positive solution to the paving conjecture (see [23] for related problems).

2.5 Extracting square submatrix with small norm : second attempt

In this section, we will show how using our main result we can answer Naor's question [55] : find an algorithm, using the Batson-Spielman-Srivastava's method [13], to prove Theorem 2.4. However, we will be able to do this only for symmetric matrices.

Proposition 2.23. *Let T be an $n \times n$ symmetric matrix with 0 diagonal. For any $\varepsilon \in (0,1)$, there exists $\sigma \subset \{1,...,n\}$ of size*

$$|\sigma| \geqslant \frac{(\sqrt{2}-1)^4 \varepsilon^2 n}{2}$$

such that

$$\left\| P_\sigma T P_\sigma^t \right\| \leqslant \varepsilon \|T\|$$

Proof. Denote $A = T + \|T\|.Id$, then A is a positive semidefinite symmetric matrix, so we may take $U = A^{\frac{1}{2}}$. First, note that since T has 0 diagonal then

$$\|Ue_i\|_2^2 = \langle Ue_i, Ue_i \rangle = \langle Ae_i, e_i \rangle = \|T\|$$

Therefore $\widetilde{U} = \frac{U}{\|T\|^{\frac{1}{2}}}$ is a standardized matrix. Moreover $\|\widetilde{U}\|^2 = 2$.

Denote $\alpha = (\sqrt{2}-1)^2$ and apply Proposition 2.19 with $(1-\alpha\varepsilon)$ to find $\sigma \subset \{1,...,n\}$ of size $\frac{\alpha^2\varepsilon^2 n}{2}$ such that

$$\frac{1-\alpha\varepsilon}{1+\alpha\varepsilon} \leqslant s_{\min}\left(\widetilde{U}_\sigma\right) \leqslant s_{\max}\left(\widetilde{U}_\sigma\right) \leqslant \frac{1+\alpha\varepsilon}{1-\alpha\varepsilon}$$

This means that

$$\left(\frac{1-\alpha\varepsilon}{1+\alpha\varepsilon}\right)^2 .Id \preceq \left(\widetilde{U}_\sigma\right)^t \cdot \left(\widetilde{U}_\sigma\right) \preceq \left(\frac{1+\alpha\varepsilon}{1-\alpha\varepsilon}\right)^2 .Id$$

Recall that $\widetilde{U}_\sigma = \widetilde{U}P_\sigma^t$ and $\widetilde{U}^t \cdot \widetilde{U} = \frac{A}{\|T\|}$. Therefore by the choice of α, we have

$$(1-\varepsilon)\|T\|.Id \preceq P_\sigma A P_\sigma^t \preceq (1+\varepsilon)\|T\|.Id,$$

which after rearrangement gives

$$-\varepsilon\|T\|.Id \preceq P_\sigma T P_\sigma^t \preceq \varepsilon\|T\|.Id$$

and finishes the proof. □

The previous result is directly related to the Kadison-Singer conjecture ([41], see also [23]). It has been proven in [7] that the Kadison-Singer conjecture is equivalent to the paving conjecture which we state here :

Conjecture. *For any $\varepsilon \in (0,1)$, there exists $p = p(\varepsilon)$ such that for any $n \in \mathbb{N}$ and any $n \times n$ matrix T with zero diagonal, there exists a partition of $\{1,...,n\}$ into p sets $\sigma_1,...,\sigma_p$ such that*

$$\forall i \leqslant p, \quad \|P_{\sigma_i} T P_{\sigma_i}^t\| \leqslant \varepsilon\|T\|$$

The difficulty lies on finding a partition whose size does not depend on the dimension n. Iterating Proposition 2.23, we obtain by a deterministic method the strongest result on the paving problem which is due to Bourgain-Tzafriri ([19], see also [86]) ; namely, every zero-diagonal matrix of size $n \times n$ can be paved with at most $O(\log(n))$ blocks. Once again, we are able to achieve this for symmetric matrices.

Proposition 2.24. *Let T be an $n \times n$ symmetric matrix with 0 diagonal. For any $\varepsilon \in (0,1)$, there exists a partition of $\{1,...,n\}$ into k subsets $\sigma_1,..,\sigma_k$ such that*

$$k \leqslant \frac{2\log(n)}{(\sqrt{2}-1)^4\varepsilon^2}$$

and for any $i \leqslant k$,

$$\left\|P_{\sigma_i} T P_{\sigma_i}^t\right\| \leqslant \varepsilon\|T\|$$

Proof. As before denote $A = T + \|T\|.Id$ and $U = A^{\frac{1}{2}}$. Note $\widetilde{U} = \frac{U}{\|T\|^{\frac{1}{2}}}$ the standardized matrix. Applying Corollary 2.22, we have a column partition for which we do on each block as we did in the previous proposition. The result follows easily. □

2.6 Application to Harmonic analysis

In this section, we will give a nice application to the problem of harmonic density in Fourier analysis. Our result improves previous work of Bourgain-Tzafriri [19] and Vershynin [89]. Let us start with some definitions :

For Λ a set of integers, the density of Λ is given by :

$$\text{dens}(\Lambda) = \lim_{n\to\infty} \frac{|\Lambda \cap \{-n, ..., n\}|}{2n}, \tag{2.9}$$

whenever the limit exists.

Definition 2.25. *Let $\mathcal{H}$ be a set of finite sets of integers. $\mathcal{H}$ is called an homogeneous system if for every $A \in \mathcal{H}$, all the subsets and translates of A belong to $\mathcal{H}$.*

Denote by $\mathbb{T}$ the circle and ν the normalized Lebesgue measure on $\mathbb{T}$. The space of twice integrable functions on $\mathbb{T}$ will be denoted by $L_2(\mathbb{T}, \nu)$. For any $f \in L_2(\mathbb{T}, \nu)$, define

$$\|f\|_{L_2(\mathbb{T})} = \left(\int_{\mathbb{T}} |f|^2 d\nu\right)^{\frac{1}{2}}$$

Let B be a subset of $\mathbb{T}$ of positive measure and define

$$\|f\|_{L_2(B)} = \left(\frac{1}{\nu(B)} \int_{\mathbb{B}} |f|^2 d\nu\right)^{\frac{1}{2}}$$

For Λ a subset of the integers, we denote the closed linear span of the characters $\{e^{i.kx}\}_{k\in\Lambda}$ in $L_2(\mathbb{T}, \nu)$ by $L_2^{\Lambda}(\mathbb{T}, \nu)$. In other terms, $L_2^{\Lambda}(\mathbb{T}, \nu)$ is the space of functions whose Fourier transforms are supported by Λ.

An interesting problem is the following :

Given a subset B of the circle $\mathbb{T}$ with $\nu(B) > 0$, can we find a subset Λ of the integers with positive density such that every function in $L_2^{\Lambda}(\mathbb{T}, \nu)$ does not vanish a.e on B ?

A positive answer was given by Bourgain-Tzafriri [19] who proved the following :

Theorem 2.26 (Bourgain-Tzafriri). *There exists a positive constant c so that, for any $B \subset \mathbb{T}$, one can find a subset Λ of the integers with $\mathrm{dens}(\Lambda) \geqslant c\nu(B)$, for which*

$$\|f\|_{L_2(B)} \geqslant c \cdot \|f\|_{L_2(\mathbb{T})}, \tag{2.10}$$

whenever $f \in L_2^\Lambda(\mathbb{T}, \nu)$.

The proof of this result makes use of the restricted invertibility principle alongside a result of Ruzsa [69] which allows to pass from large finite sets to an infinite set of large density.

Theorem 2.27 (Ruzsa). *Given an arbitrary homogeneous system $\mathcal{H}$, there exists a set of integers Λ such that its finite subsets all belong to $\mathcal{H}$ and*

$$\mathrm{dens}(\Lambda) = d(\mathcal{H}),$$

where $d(\mathcal{H})$ is given by

$$d(\mathcal{H}) = \lim_{n\to\infty} \max_{A\in\mathcal{H}} \frac{|A \cap \{1, .., n\}|}{n}.$$

Remark 2.28. *In [69], Ruzsa proved the previous statement with the definition of density (2.9) replaced by the one-sided density $\lim_{n\to\infty} \frac{|\Lambda\cap\{1,..,n\}|}{n}$. In [89], Vershynin noticed that a slight modification of Ruzsa's argument yields the result stated in Theorem 2.27.*

Following the same strategy as in [19], Vershynin [89] added a non trivial upper bound to (2.10). This follows from the use of Theorem 2.1. Therefore, making use of Theorem 2.2, we improve the previous results and show the following :

Theorem 2.29. *Let B be a subset of $\mathbb{T}$ of positive measure. For any $\varepsilon \in (0, 1)$, there exists a set of integers Λ with density $\mathrm{dens}(\Lambda) \geqslant (1-\varepsilon)^2\nu(B)$ such that for any $f \in L_2^\Lambda(\mathbb{T}, \nu)$, we have*

$$\frac{\varepsilon}{2-\varepsilon} \|f\|_{L_2(\mathbb{T})} \leqslant \|f\|_{L_2(B)} \leqslant \frac{2-\varepsilon}{\varepsilon} \|f\|_{L_2(\mathbb{T})}$$

Proof. Define T a linear operator on $L_2(\mathbb{T}, \nu)$ as follows

$$T(f) = \frac{1}{\sqrt{\nu(B)}} f \cdot \mathbf{1}_B; \quad f \in L_2(\mathbb{T}, \nu).$$

One can easily check that $\|T\| = \nu(B)^{-\frac{1}{2}}$ and $\|T(e^{i.kx})\| = 1$ for all $k \in \mathbb{Z}$. For every positive integer n, taking U, the matrix whose columns are the coordinates of $\left(T(e^{i.kx})\right)_{k\leqslant n}$ in some basis, and applying Theorem 2.2 we get

$$\sigma_n \subset \{1, ..., n\}$$

of size

$$|\sigma_n| \geqslant (1-\varepsilon)^2 \frac{n}{\|U\|^2} \geqslant (1-\varepsilon)^2 \frac{n}{\|T\|^2} = (1-\varepsilon)^2 n \cdot \nu(B)$$

such that for any $f \in L_2^{\sigma_n}(\mathbb{T}, \nu)$ we have

$$\frac{\varepsilon}{2-\varepsilon} \|f\|_{L_2(\mathbb{T})} \leqslant \|T(f)\|_{L_2(\mathbb{T})} \leqslant \frac{2-\varepsilon}{\varepsilon} \|f\|_{L_2(\mathbb{T})} .$$

Since $\|T(f)\|_{L_2(\mathbb{T})} = \|f\|_{L_2(B)}$ then the previous equation yields the following for any n and any $f \in L_2^{\sigma_n}(\mathbb{T}, \nu)$

$$\frac{\varepsilon}{2-\varepsilon} \|f\|_{L_2(\mathbb{T})} \leqslant \|f\|_{L_2(B)} \leqslant \frac{2-\varepsilon}{\varepsilon} \|f\|_{L_2(\mathbb{T})} . \tag{2.11}$$

Define $\mathcal{H}$ the family of all finite subsets σ of the integers such that (2.11) is satisfied for any $f \in L_2^{\sigma}(\mathbb{T}, \nu)$. It is easy to check that $\mathcal{H}$ is an homogeneous system. Moreover, since $\sigma_n \in \mathcal{H}$ for any n then

$$d(\mathcal{H}) \geqslant \limsup_{n\to\infty} \frac{|\sigma_n|}{n} \geqslant (1-\varepsilon)^2 \nu(B).$$

Applying Theorem 2.27, we get a set of integers Λ with $\text{dens}(\Lambda) \geqslant (1-\varepsilon)^2\nu(B)$ such that all its finite subsets belong to $\mathcal{H}$. This completes the proof in view of the definition of $\mathcal{H}$. □

In the regime where ε is close to one, we get the following :

Corollary 2.30. *Let B be a subset of $\mathbb{T}$ of positive measure. For any $\varepsilon \in (0,1)$, there exists a set of integers Λ with density* $\text{dens}(\Lambda) \geqslant \frac{\varepsilon^2}{9}\nu(B)$ *such that for any $f \in L_2^{\Lambda}(\mathbb{T}, \nu)$, we have*

$$(1-\varepsilon) \|f\|_{L_2(\mathbb{T})} \leqslant \|f\|_{L_2(B)} \leqslant (1+\varepsilon) \|f\|_{L_2(\mathbb{T})}$$

Chapitre 3

Banach-Mazur distance to the cube

3.1 Introduction

Measuring distance between objects is an important tool towards identifying these ones. In our context, the local theory of Banach spaces, two spaces are identified if they are isometric. One can see an n-dimensional Banach space as $\mathbb{R}^n$ equipped with a norm. It is well known that all norms on $\mathbb{R}^n$ are equivalent ; this means that for $|\cdot|$ and $\|\cdot\|$ two different norms on $\mathbb{R}^n$, there exist positive constants α and β such that

$$\beta|\cdot| \leqslant \|\cdot\| \leqslant \alpha|\cdot|.$$

The constant $\gamma = \frac{\alpha}{\beta}$ will measure the "degree" of equivalence of these two norms. If the constant of equivalence γ is one, then the two norms are just homothetic and can be identified. Of course γ can depend on the dimension n and the aim is to trace this dependence. Since we are interested in "big" dimensions, if γ does not depend on the dimension then it is considered small and the two norms are well identified. Defining the distance between two Banach spaces as the constant of equivalence between their norms is known as the geometric distance. In this chapter, we will discuss the Banach-Mazur distance which is a smaller quantity than the geometric distance. Indeed, to calculate the Banach-Mazur distance we allow some linear transformation on the original spaces then calculate the geometric distance between the obtained spaces. The aim of this chapter is to estimate the distance of an n-dimensional Banach space to the n-dimensional cube l_∞^n. To this end, we discuss the proportional Dvoretzky-Rogers factorization which we can derive using our results from previous chapters and then use this to estimate the distance to the cube.

Let $\mathbb{BM}_n$ denote the space of all n-dimensional normed spaces X, known as the Banach-Mazur compactum. If X, Y are in $\mathbb{BM}_n$, we define the Banach-Mazur distance between X and Y as follows :

$$d(X,Y) = \inf\{\|T\| \cdot \|T^{-1}\| \mid \; T \text{ is an isomorphism between X and Y } \}$$

Remark 3.1. *For K, L two symmetric convex bodies in $\mathbb{R}^n$, one can define the Banach-Mazur distance between K and L as*

$$d(K,L) = \inf\{\alpha/\beta \mid \; \beta L \subset T(K) \subset \alpha L\}$$

One can easily check that this distance is coherent with the previous one as $d(X,Y) = d(B_X, B_Y)$.

Remark 3.2. *It is easy to check that d is multiplicative and that $d(X,Y) = 1$ if and only if X and Y are isometric. Therefore $\log(d)$ is a distance over the quotient of $\mathbb{BM}_n$ obtained by identifying isometric spaces.*

3.2 Proportional Dvoretzky-Rogers factorization

By the classical Dvoretzky-Rogers lemma [28], it is proven that if X is an n-dimensional Banach space then there exist $x_1, ..., x_k \in X$ with $k = \sqrt{n}$ such that for all scalars $(a_j)_{j\leqslant k}$

$$\max_{j\leqslant k} |a_j| \leqslant \left\| \sum_{j\leqslant k} a_j x_j \right\|_X \leqslant c \left(\sum_{j\leqslant k} a_j^2 \right)^{\frac{1}{2}},$$

where c is a universal constant. Bourgain-Szarek [18] proved that the previous statement holds for k proportional to n, and called the result "the proportional Dvoretzky-Rogers factorization" :

Theorem 3.3 (Proportional Dvoretzky-Rogers factorization)**.** *Let X be an n-dimensional Banach space. For any $\varepsilon \in (0,1)$, there exist $x_1, ..., x_k \in X$ with $k \geqslant (1-\varepsilon)n$ such that for all scalars $(a_j)_{j\leqslant k}$*

$$\max_{j\leqslant k} |a_j| \leqslant \left\| \sum_{j\leqslant k} a_j x_j \right\|_X \leqslant c(\varepsilon) \left(\sum_{j\leqslant k} a_j^2 \right)^{\frac{1}{2}},$$

where $c(\varepsilon)$ is a constant depending on ε. Equivalently, the identity operator $i_{2,\infty} : l_2^k \longrightarrow l_\infty^k$ can *be written* $i_{2,\infty} = \alpha \circ \beta$ *with* $\beta : l_2^k \longrightarrow X, \alpha : X \longrightarrow l_\infty^k$ *and* $\|\alpha\| \cdot \|\beta\| \leqslant c(\varepsilon)$.

Finding the right dependence on ε is an important problem and the optimal result is not known yet. In [80], Szarek showed that one cannot hope for a dependence better than $c\varepsilon^{-\frac{1}{10}}$. Szarek-Talagrand [82] proved that the previous result holds with $c(\varepsilon) = c\varepsilon^{-2}$ and in [31] and [32] Giannopoulos improved the dependence to get $c\varepsilon^{-\frac{3}{2}}$ and $c\varepsilon^{-1}$. In all these results, a factorization for the identity operator $i_{1,2} : l_1^n \longrightarrow l_2^n$ was proven and by duality the factorization for $i_{2,\infty}$ was deduced. The previous proofs used some geometric results, technical combinatorics and Grothendieck's factorization theorem ([27], [61]). Here we present a direct proof using Theorem 1.4 which allows us to recover the best known dependence on ε and improve the universal constant involved.

As we have seen in Chapter 1, the set of n-dimensional Banach spaces can be identified with the set of symmetric convex bodies in $\mathbb{R}^n$. Therefore, Theorem 3.3 can be formulated with symmetric convex bodies. In [50], Litvak and Tomczak-Jaegermann proved a nonsymmetric version of the proportional Dvoretzky-Rogers factorization :

Theorem 3.4 (Litvak-Tomczak-Jaegermann)**.** *Let* $K \subset \mathbb{R}^n$ *be a convex body, such that* B_2^n *is the ellipsoid of minimal volume containing* K*. Let* $\varepsilon \in (0,1)$ *and set* $k = [(1-\varepsilon)n]$*. There exist vectors* $y_1, y_2, ..., y_k$ *in* K*, and an orthogonal projection* P *in* $\mathbb{R}^n$ *with rank*$(P) \geqslant k$ *such that for all scalars* $t_1, ..., t_k$

$$c\varepsilon^3 \left(\sum_{j=1}^k |t_j|^2 \right)^{\frac{1}{2}} \leqslant \left\| \sum_{j=1}^k t_j P y_j \right\|_{PK} \leqslant \frac{6}{\varepsilon} \sum_{j=1}^k |t_j|,$$

where $c > 0$ *is a universal constant.*

Using again Theorem 1.4 combined with some tools developped in [50] and [18], we will be able to improve the dependence on ε in the previous statement.

3.2.1 The symmetric case

Let us start with the original proportional Dvoretzky-Rogers factorization. We will prove the following :

Theorem 3.5. *Let* X *be an* n*-dimensional Banach space. For any* $\varepsilon \in (0,1)$*, there exist* $x_1, ..., x_k \in X$ *with* $k \geqslant [(1-\varepsilon)^2 n]$ *such that for all scalars* $(a_j)_{j \leqslant k}$

$$\varepsilon \left(\sum_{j \leqslant k} a_j^2 \right)^{\frac{1}{2}} \leqslant \left\| \sum_{j \leqslant k} a_j x_j \right\|_X \leqslant \sum_{j \leqslant k} |a_j|$$

Equivalently, the identity operator $i_{1,2} : l_1^k \longrightarrow l_2^k$ *can be written as* $i_{1,2} = \alpha \circ \beta$*, where* $\beta : l_1^k \longrightarrow X$*,* $\alpha : X \longrightarrow l_2^k$ *and* $\|\alpha\| \cdot \|\beta\| \leqslant \varepsilon^{-1}$*.*

Proof. Without loss of generality, we may assume that $X = (\mathbb{R}^n, \|\cdot\|_X)$ and B_2^n is the ellipsoid of minimal volume containing B_X. By John's theorem [40] there exist $x_1, ..., x_m$ contact points of B_X with B_2^n ($\|x_j\|_X = \|x_j\|_{X^*} = \|x_j\|_2 = 1$) and positive scalars $c_1, ..., c_m$ such that

$$Id = \sum_{j \leqslant m} c_j x_j x_j^t \quad \text{and} \quad \|\cdot\|_2 \leqslant \|\cdot\|_X$$

Let $U = \left(\sqrt{c_1}x_1, ..., \sqrt{c_m}x_m\right)$ be the $n \times m$ rectangular matrix whose columns are $\sqrt{c_j}x_j$ and denote $D = diag(\sqrt{c_1}, ..., \sqrt{c_m})$ the $m \times m$ diagonal matrix with $\sqrt{c_j}$ on its diagonal. Clearly, $UU^t = Id$ and $\|U\|_{\mathrm{HS}} = \|D\|_{\mathrm{HS}} = \sqrt{n}$.
Let $\varepsilon < 1$, applying Theorem 1.4 to U and D, we find $\sigma \subset \{1, ..., m\}$ such that

$$k = |\sigma| \geqslant \left[(1-\varepsilon)^2 n\right]$$

and for all $a = (a_j)_{j \in \sigma}$

$$\left\|U_\sigma D_\sigma^{-1} a\right\|_2 = \left\|\sum_{j \in \sigma} a_j x_j\right\|_2 \geqslant \varepsilon \left(\sum_{j \in \sigma} |a_j|^2\right)^{\frac{1}{2}} \tag{3.1}$$

To simplify the notations, we may assume that $\sigma = \{1, \ldots, k\}$. Denote P the orthogonal projection of X onto $Y = \operatorname{span}\{(x_j)_{j \leqslant k}\}$. Now note that (3.1) guarantees that the $(x_j)_{j \leqslant k}$ are linearly independent and therefore that P is of rank k. Define T and β as follows :

$$\begin{array}{rcccl} \beta & : & l_1^k & \to & X \\ & & e_j & \mapsto & x_j \quad \text{for } j \leqslant k \end{array} \qquad \text{and} \qquad \begin{array}{rcccl} T & : & Y & \to & l_2^k \\ & & x_j & \mapsto & e_j \quad \text{for } j \leqslant k \end{array}$$

and write $\alpha = TP$. For $a = (a_j)_{j \leqslant k} \in \mathbb{R}^k$, by the triangle inequality we have

$$\|\beta(a)\|_X = \left\|\sum_{j \leqslant k} a_j x_j\right\|_X \leqslant \sum_{j \leqslant k} |a_j| \cdot \|x_j\|_X = \sum_{j \leqslant k} |a_j|,$$

and therefore $\|\beta\| \leqslant 1$. Now let $x \in X$, then $Px \in Y$ and one can write $Px = \sum_{j \leqslant k} a_j x_j$. Using

(3.1) we get

$$\|\alpha(x)\|_2 = \left(\sum_{j\leqslant k} a_j^2\right)^{\frac{1}{2}} \leqslant \frac{1}{\varepsilon}\left\|\sum_{j\leqslant k} a_j x_j\right\|_2 = \frac{1}{\varepsilon}\|Px\|_2 \leqslant \frac{1}{\varepsilon}\|x\|_2 \leqslant \frac{1}{\varepsilon}\|x\|_X,$$

and therefore $\|\alpha\| \leqslant \varepsilon^{-1}$, which finishes the proof. □

Remark 3.6. *Denoting $\mathcal{F}$ the ellipsoid of minimal volume containing B_X, we proved the existence of vectors $x_1, .., x_k$ with $k \geqslant [(1-\varepsilon)^2 n]$ such that for all scalars $(a_j)_{j\leqslant k}$*

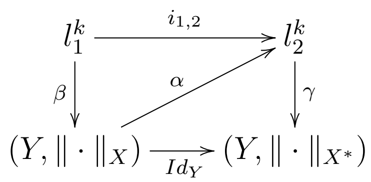
$$\varepsilon\left(\sum_{j\leqslant k} a_j^2\right)^{\frac{1}{2}} \leqslant \left\|\sum_{j\leqslant k} a_j x_j\right\|_{\mathcal{F}} \leqslant \left\|\sum_{j\leqslant k} a_j x_j\right\|_X \leqslant \sum_{j\leqslant k}|a_j| \tag{3.2}$$

As a direct application of the previous result, we have

Corollary 3.7. *Let X be an n-dimensional Banach space. For any $\varepsilon \in (0,1)$, there exists Y a subspace of X of dimension $k \geqslant [(1-\varepsilon)^2 n]$ such that $d(Y, l_1^k) \leqslant \frac{\sqrt{n}}{\varepsilon}$.*

Instead of using Theorem 1.4, we could have used proposition 2.15 but we lose a constant factor of $\frac{1}{2}$ in the estimate. Nevertheless, we can get the following factorization :

Theorem 3.8. *Let $X = (\mathbb{R}^n, \|\cdot\|)$ where $\|\cdot\|$ is a norm on $\mathbb{R}^n$ such that B_2^n is the ellipsoid of minimal volume containing B_X. For any $\varepsilon \in (0,1)$, there exists $Y \subset \mathbb{R}^n$ a subspace of dimension $k \geqslant \lfloor(1-\varepsilon)^2 n\rfloor$ such that the following diagram commutes*



i.e $i_{1,2} = \alpha \circ \beta$ and $Id_Y = \gamma \circ \alpha$. Moreover $\|\beta\| \leqslant 1$, $\|\alpha\| \leqslant \frac{2-\varepsilon}{\varepsilon}$ and $\|\gamma\| \leqslant \frac{2-\varepsilon}{\varepsilon}$.

Proof. Let $\varepsilon < 1$, we may start applying Proposition 2.15 to find $x_1, ..., x_k$ contact points of B_X with B_2^n such that

$$k \geqslant \left\lfloor(1-\varepsilon)^2 n\right\rfloor$$

and for all scalars $(a_j)_{j\leqslant k}$,

$$\frac{\varepsilon}{2-\varepsilon}\left(\sum_{j\leqslant k} a_j^2\right)^{\frac{1}{2}} \leqslant \left\|\sum_{j\leqslant k} a_j x_j\right\|_2 \leqslant \frac{2-\varepsilon}{\varepsilon}\left(\sum_{j\leqslant k} a_j^2\right)^{\frac{1}{2}}. \tag{3.3}$$

Define Y =span$\{(x_j)_{j\leqslant k}\}$. Since $x_1, ..., x_k$ are linearly independent then Y is of dimension k. We may now define α, β, γ as follows :

$$\begin{array}{lllll} \beta & : & l_1^k & \to & Y \\ & & e_j & \mapsto & x_j \quad \forall j \leqslant k \end{array} \qquad \begin{array}{lllll} \alpha & : & Y & \to & l_2^k \\ & & x_j & \mapsto & e_j \quad \forall j \leqslant k \end{array} \qquad \begin{array}{lllll} \gamma & : & l_2^k & \to & Y \\ & & e_j & \mapsto & x_j \quad \forall j \leqslant k \end{array}$$

For $a = (a_j)_{j\leqslant k} \in \mathbb{R}^k$, by the triangle inequality we have

$$\|\beta(a)\|_X = \left\| \sum_{j\leqslant k} a_j x_j \right\|_X \leqslant \sum_{j\leqslant k} |a_j| \cdot \|x_j\|_X = \|a\|_1,$$

and therefore $\|\beta\| \leqslant 1$. Since $B_X \subset B_2^n$ then

$$\|\cdot\|_{X^*} \leqslant \|\cdot\|_2 \leqslant \|\cdot\|_X$$

Therefore, using (3.3) we have for $a = (a_j)_{j\leqslant k} \in \mathbb{R}^k$

$$\|\gamma(a)\|_{X^*} \leqslant \|\gamma(a)\|_2 = \left\| \sum_{j\leqslant k} a_j x_j \right\|_2 \leqslant \frac{2-\varepsilon}{\varepsilon} \left(\sum_{j\leqslant k} a_j^2 \right)^{\frac{1}{2}},$$

which means that $\|\gamma\| \leqslant \frac{2-\varepsilon}{\varepsilon}$.
Let $x \in Y$ then $x = \sum_{j\leqslant k} a_j x_j$ for some $a = (a_j)_{j\leqslant k}$ and therefore using (3.3) we have

$$\|\alpha(x)\|_2 = \left(\sum_{j\leqslant k} a_j^2 \right)^{\frac{1}{2}} \leqslant \frac{2-\varepsilon}{\varepsilon} \left\| \sum_{j\leqslant k} a_j x_j \right\|_2 \leqslant \frac{2-\varepsilon}{\varepsilon} \left\| \sum_{j\leqslant k} a_j x_j \right\|_X,$$

which means that $\|\alpha\| \leqslant \frac{2-\varepsilon}{\varepsilon}$. □

3.2.2 The nonsymmetric case

Let us now turn to the non symmetric version of Theorem 3.5. We will prove the following :

Theorem 3.9. *Let $K \subset \mathbb{R}^n$ be a convex body, such that B_2^n is the ellipsoid of minimal volume containing K. For any $\varepsilon \in (0,1)$, there exist $x_1, ..., x_k$ with $k \geqslant (1-\varepsilon)n$ contact points and P an orthogonal projection of rank $\geqslant k$ such that for all $(a_j)_{j\leqslant k}$*

$$\frac{\varepsilon^2}{16} \left(\sum_{j=1}^k |a_j|^2 \right)^{\frac{1}{2}} \leqslant \left\| \sum_{j=1}^k a_j P x_j \right\|_{PK} \leqslant \frac{4}{\varepsilon} \sum_{j=1}^k |a_j|$$

Proof. By John's Theorem [40], we get an identity decomposition in $\mathbb{R}^n$

$$Id = \sum_{j \leqslant m} c_j x_j x_j^t \quad \text{and} \quad \sum_{j \leqslant m} c_j x_j = 0,$$

where $x_1, ..., x_m$ are contact points of K and B_2^n and $(c_j)_{j \leqslant m}$ positive scalars. Note that we will not use the second assertion i.e the fact that $\sum_{j \leqslant m} c_j x_j = 0$.

As before, take $U = \left(\sqrt{c_1}x_1, ..., \sqrt{c_m}x_m\right)$ the $n \times m$ rectangular matrix whose columns are $\sqrt{c_j}x_j$. Denote $D = diag(\sqrt{c_1}, ..., \sqrt{c_m})$ the $m \times m$ diagonal matrix with $\sqrt{c_j}$ on its diagonal. Applying Theorem 1.4 to U and D with parameter $\frac{\varepsilon}{4}$, we find $\sigma_1 \subset \{1, ..., m\}$ such that

$$s = |\sigma_1| \geqslant \left(1 - \frac{\varepsilon}{4}\right)^2 n \geqslant (1 - \frac{\varepsilon}{2})n,$$

and for all scalars $a = (a_j)_{j \in \sigma_1}$

$$\left\|U_{\sigma_1} D_{\sigma_1}^{-1} a\right\|_2 = \left\|\sum_{j \in \sigma_1} a_j x_j\right\|_2 \geqslant \frac{\varepsilon}{4} \left(\sum_{j \in \sigma_1} |a_j|^2\right)^{\frac{1}{2}}. \tag{3.4}$$

Define $Y = \text{span}\{x_j\}_{j \in \sigma_1}$. We will now use the argument in [50] to construct the projection P. First partition σ_1 into $\left[\frac{\varepsilon}{2}s\right]$ disjoint subsets A_l of equal size. Clearly

$$|A_l| \leqslant \left[\frac{s}{[\frac{\varepsilon}{2}s]}\right] + 1 \leqslant \left[\frac{2}{\varepsilon} \cdot \frac{\frac{\varepsilon}{2}s}{[\frac{\varepsilon}{2}s]}\right] + 1 \leqslant \left[\frac{4}{\varepsilon}\right] + 1$$

Let $z_l = \sum_{i \in A_l} x_i$ and take $P : Y \longrightarrow Y$ the orthogonal projection onto $\text{span}\{z_l\}^{\perp}$. For every l, we have $Pz_l = 0$ so that for $j \in A_l$ we can write

$$-Px_j = \sum_{i \in A_l, i \neq j} Px_i = (|A_l| - 1) \cdot \frac{1}{|A_l| - 1} \sum_{i \in A_l, i \neq j} Px_i$$

We deduce that for every l and every $j \in A_l$, we have $-Px_j \in (|A_l| - 1)\, PK \subset \frac{4}{\varepsilon} PK$.

Let $T : \mathbb{R}^{|\sigma_1|} \longrightarrow Y$ be a linear operator defined by $Te_j = x_j$ for all $j \in \sigma_1$, where $(e_j)_{j \in \sigma_1}$ denotes the canonical basis of $\mathbb{R}^{|\sigma_1|}$ and Y is equipped with the euclidean norm. Since $(x_j)_{j \in \sigma_1}$ are linearly independent, T is an isomorphism. Moreover, by (3.4), we have $\|T^{-1}\| \leqslant \frac{4}{\varepsilon}$. Take $P' = T^{-1}PT$ and P'' the orthogonal projection onto $(\text{Ker}P')^{\perp}$. It is easy to check that

$P''P' = P''$ and

$$k = \text{rank}P'' = \text{rank}P \geqslant \left(1 - \frac{\varepsilon}{2}\right) s \geqslant (1 - \varepsilon)\, n$$

For all scalars $(a_j)_{j\in\sigma_1}$,

$$\begin{aligned}\left\| \sum_{j\in\sigma_1} a_j P x_j \right\|_2 &= \left\| \sum_{j\in\sigma_1} a_j P T e_j \right\|_2 \\ &= \left\| \sum_{j\in\sigma_1} T\left(a_j P' e_j\right) \right\|_2 \\ &\geqslant \frac{1}{\|T^{-1}\|} \cdot \left\| \sum_{j\in\sigma_1} a_j P' e_j \right\|_2 \\ &\geqslant \frac{\varepsilon}{4} \cdot \left\| \sum_{j\in\sigma_1} a_j P'' e_j \right\|_2\end{aligned}$$

Now take U' the $s \times s$ matrix whose columns are $(P''e_j)_{j\in\sigma_1}$. Apply Theorem 1.4 with U' and Id as diagonal matrix and $\frac{\varepsilon}{4}$ as parameter, then there exists $\sigma \subset \sigma_1$ of size

$$|\sigma| \geqslant \left(1 - \frac{\varepsilon}{4}\right)^2 s \geqslant (1 - \varepsilon)n$$

such that for all scalars $(a_j)_{j\in\sigma}$,

$$\left\| \sum_{j\in\sigma} a_j P'' e_j \right\|_2 \geqslant \frac{\varepsilon}{4} \left(\sum_{j\in\sigma} |a_j|^2 \right)^{\frac{1}{2}}.$$

This gives us the following

$$\left\| \sum_{j\in\sigma} a_j P x_j \right\|_2 \geqslant \frac{\varepsilon}{4} \cdot \left\| \sum_{j\in\sigma} a_j P'' e_j \right\|_2 \geqslant \frac{\varepsilon^2}{16} \left(\sum_{j\in\sigma} |a_j|^2 \right)^{\frac{1}{2}}.$$

On the other hand, since $K \subset B_2^n$ then

$$\left\| \sum_{j\in\sigma} a_j P x_j \right\|_2 = \left\| \sum_{j\in\sigma} a_j P x_j \right\|_{PB_2^n} \leqslant \left\| \sum_{j\in\sigma} a_j P x_j \right\|_{PK}$$

Denoting $A = -PK \cap PK$ which is a centrally symmetric convex body, and using the fact that $-Px_j \in \frac{4}{\varepsilon} PK$ alongside the triangle inequality, one can write

$$\left\|\sum_{j\in\sigma} a_j Px_j\right\|_A \leqslant \frac{4}{\varepsilon}\sum_{j\in\sigma}|a_j|.$$

Finally, we have

$$\frac{\varepsilon^2}{16}\left(\sum_{j\in\sigma}|a_j|^2\right)^{\frac{1}{2}} \leqslant \left\|\sum_{j\in\sigma} a_j Px_j\right\|_2 \leqslant \left\|\sum_{j\in\sigma} a_j Px_j\right\|_{PK} \leqslant \left\|\sum_{j\in\sigma} a_j Px_j\right\|_A \leqslant \frac{4}{\varepsilon}\sum_{j\in\sigma}|a_j|$$

□

One can interpret the previous result geometrically as follows :

Corollary 3.10. *Let $K \subset \mathbb{R}^n$ be a convex body such that B_2^n is the ellipsoid of minimal volume containing K. For any $\varepsilon \in (0,1)$, there exists P an orthogonal projection of rank $k \geqslant [(1-\varepsilon)n]$ such that*

$$\frac{\varepsilon}{4}B_1^k \subset PK \subset \frac{16}{\varepsilon^2}B_2^k.$$

Moreover, $d(PK, B_1^k) \leqslant \frac{64\sqrt{n}}{\varepsilon^3}$.

By duality, this means that there exists a subspace $E \subset \mathbb{R}^n$ of dimension $k \geqslant [(1-\varepsilon)n]$ such that

$$\frac{\varepsilon^2}{16}B_2^k \subset K \cap E \subset \frac{4}{\varepsilon}B_\infty^k.$$

Moreover, $d(K\cap E, B_\infty^k) \leqslant \frac{64\sqrt{n}}{\varepsilon^3}$.

3.3 Estimate of the Banach-Mazur distance to the Cube

In [18], Bourgain-Szarek showed how to estimate the Banach-Mazur distance to the cube once a proportional Dvoretzky-Rogers factorization is proven. This technique was again used in [82] and [31]. Since we are able to obtain a proportional Dvoretzky-Rogers factorization with a better constant, then using the same argument we will recover the best known asymptotic for the Banach-Mazur distance to the cube and improve the constants involved. Let us start defining

$$R_\infty^n = \max\left\{d(X, l_\infty^n)/\ X \in \mathbb{BM}_n\right\}$$

Similarly one can define R_1^n, and since the Banach-Mazur distance is invariant by duality then $R_1^n = R_\infty^n$. Since the diameter of $\mathbb{BM}_n$ is less than n then a trivial estimate would be $R_\infty^n \leqslant n$. In [80], Szarek showed the existence of an n-dimensional Banach space X such that

$d(X, l_\infty^n) \geqslant c\sqrt{n}\log(n)$. Bourgain-Szarek proved in [18] that $R_\infty^n \leqslant o(n)$ while Szarek-Talagrand [82] and Giannopoulos [31] improved this upper bound to $cn^{\frac{7}{8}}$ and $cn^{\frac{5}{6}}$ respectively. Here, we will prove the following estimate :

Theorem 3.11. *Let X be an n-dimensional Banach space. Then*

$$d(X, l_1^n) \leqslant 2^{\frac{4}{3}}\sqrt{n} \cdot d(X, l_2^n)^{\frac{2}{3}}.$$

proof. We denote $d_X = d(X, l_2^n)$. In order to bound $d(X, l_1^n)$, we need to define an isomorphism $T : l_1^n \longrightarrow X$ and estimate $\|T\| \cdot \|T^{-1}\|$. A natural way is to find a basis of X and then define T the operator which sends the canonical basis of $\mathbb{R}^n$ to this basis of X. The main idea is to find a "large" subspace Y of X which is "not too far" from l_1 (actually more is needed), then complement the basis of Y to obtain a basis of X. Finding the "large" subspace is the heart of the method and is basically given by the proportional Dvoretzky-Rogers factorization. The proof is mainly divided in four steps :

-First step : Place B_X into a "good" position and choose the right euclidean structure. Since the Banach-Mazur distance is invariant under linear transformation, we may change the position of B_X. Therefore without loss of generality we may assume that $X = (\mathbb{R}^n, \|\cdot\|_X)$ and B_2^n is the ellipsoid of minimal volume containing B_X. Denote also $\mathcal{E}$ the distance ellipsoid i.e

$$\frac{1}{d_X}\mathcal{E} \subset B_X \subset \mathcal{E} \tag{3.5}$$

The ellipsoid $\mathcal{E}$ can be defined as

$$\mathcal{E} = \left\{ x \in \mathbb{R}^n / \sum_{j=1}^{n} \alpha_i^2 \langle x, v_j \rangle^2 \leqslant 1 \right\},$$

where v_j is an orthonormal basis (in the standard sense) of $\mathbb{R}^n$ and α_j positive scalars. To take into consideration the two euclidean structures, we will define the following ellipsoid

$$\mathcal{E}_1 = \left\{ x \in \mathbb{R}^n / \sum_{j=1}^{n} \frac{1}{2}\left(1 + \alpha_i^2\right) \langle x, v_j \rangle^2 \leqslant 1 \right\}.$$

It is easy to check that

$$B_2^n \cap \mathcal{E} \subset \mathcal{E}_1 \subset \sqrt{2} B_2^n \cap \mathcal{E} \tag{3.6}$$

Therefore

$$\frac{1}{\sqrt{2d_X}}\mathcal{E}_1 \subset B_X \subset \mathcal{E}_1 \tag{3.7}$$

-Second step : Let $\varepsilon > 0$ and set $k = (1-2\varepsilon)n$. Apply Remark 3.6 to find $x_1, ..., x_k$ in X such that for all scalars $(a_j)_{j\leqslant k}$

$$\varepsilon\left(\sum_{j\leqslant k} a_j^2\right)^{\frac{1}{2}} \leqslant \left\|\sum_{j\leqslant k} a_j x_j\right\|_2 \leqslant \left\|\sum_{j\leqslant k} a_j x_j\right\|_X \leqslant \sum_{j\leqslant k} |a_j| \tag{3.8}$$

Note that $(x_j)_{j\leqslant k}$ are linearly independent and are a good candidate to be part of the basis of X.

-Third step : To form a basis of X, we simply take $y_{k+1}, .., y_n$ an orthogonal basis in the $\mathcal{E}_1$-sense of $\mathrm{span}\{(x_j)_{j\leqslant k}\}^{\perp}$ (where the $\perp$ is in the $\mathcal{E}_1$-sense) such that $\|y_j\|_{\mathcal{E}_1} = \frac{1}{\sqrt{2d_X}}$. By (3.7), we have

$$\forall j > k, \quad \|y_j\|_X \leqslant 1$$

-Fourth step : Define $T : l_1^k \longrightarrow X$ by $T(e_j) = x_j$ if $j \leqslant k$ and $T(e_j) = y_j$ if $j > k$. Let $a = (a_j)_{j\leqslant n} \in \mathbb{R}^n$ and write

$$Ta = \sum_{j=1}^{k} a_j x_j + \sum_{j=k+1}^{n} a_j y_j.$$

Then using the triangle inequality and (3.7), one can write

$$\|a\|_1 = \sum_{j\leqslant k} |a_j| + \sum_{j>k} |a_j| \geqslant \left\|\sum_{j\leqslant k} a_j x_j + \sum_{j>k} a_j y_j\right\|_X \geqslant \left\|\sum_{j\leqslant k} a_j x_j + \sum_{j>k} a_j y_j\right\|_{\mathcal{E}_1}.$$

We also have

$$
\begin{aligned}
\|Ta\|_{\mathcal{E}_1} &\geqslant \left[\Big\|\sum_{j\leqslant k} a_j x_j\Big\|_{\mathcal{E}_1}^2 + \Big\|\sum_{j>k} a_j y_j\Big\|_{\mathcal{E}_1}^2\right]^{\frac{1}{2}} && \text{by orthogonality}\\
&\geqslant \left[\frac{1}{2}\Big\|\sum_{j\leqslant k} a_j x_j\Big\|_{2}^2 + \Big\|\sum_{j>k} a_j y_j\Big\|_{\mathcal{E}_1}^2\right]^{\frac{1}{2}} && \text{by (3.6)}\\
&\geqslant \left[\frac{1}{2}\varepsilon^2\sum_{j\leqslant k} a_j^2 + \sum_{j>k} a_j^2\|y_j\|_{\mathcal{E}_1}^2\right]^{\frac{1}{2}} && \text{by (3.8)}\\
&\geqslant \left[\frac{\varepsilon^2}{2n}\Big(\sum_{j\leqslant k}|a_j|\Big)^2 + \frac{1}{2d_X^2(n-k)}\Big(\sum_{j>k}|a_j|\Big)^2\right]^{\frac{1}{2}} && \text{by Cauchy-Shwarz}\\
&\geqslant \left[\frac{\varepsilon^2}{2n}\Big(\sum_{j\leqslant k}|a_j|\Big)^2 + \frac{1}{4\varepsilon n d_X^2}\Big(\sum_{j>k}|a_j|\Big)^2\right]^{\frac{1}{2}}\\
&\geqslant \frac{1}{2}\left[\frac{\varepsilon}{\sqrt{n}}\sum_{j\leqslant k}|a_j| + \frac{1}{d_X\sqrt{2\varepsilon n}}\sum_{j>k}|a_j|\right]\\
&\geqslant \frac{1}{2^{\frac{4}{3}}\sqrt{n}d_X^{\frac{2}{3}}}\sum_{j=1}^{n}|a_j| && \text{taking } \varepsilon=(\sqrt{2}d_X)^{-\frac{2}{3}}.
\end{aligned}
$$

As a conclusion,

$$\frac{1}{2^{\frac{4}{3}}\sqrt{n}d_X^{\frac{2}{3}}}\|a\|_1 \leqslant \|Ta\|_X \leqslant \|a\|_1$$

and therefore $d(X, l_1^n) \leqslant 2^{\frac{4}{3}}\sqrt{n}d_X^{\frac{2}{3}}$ for all $X \in \mathbb{BM}_n$. □

Using the same procedure and working only with one ellipsoid $\mathcal{F}$, the ellipsoid of minimal volume containing B_X, and noting that by John's theorem [40] $\frac{1}{\sqrt{n}}\mathcal{F} \subset B_X \subset \mathcal{F}$, we get the following

Theorem 3.12. $R_1^n = R_\infty^n \leqslant (2n)^{\frac{5}{6}}$.

Remark 3.13. Here we are interested in high dimensional results ; this is why the constant is not that important. If we want an estimate for "small" dimensions, then the value of the constant becomes important. In [31], Giannopoulos proved that $R_\infty^n \leqslant cn^{\frac{5}{6}}$ with $c = \frac{2^{\frac{7}{6}}}{(\sqrt{2}-1)^{\frac{1}{3}}} \sim 3,0116$, and thus his result becomes nontrivial when the dimension is larger than 747. On the other hand, our result becomes nontrivial whenever the dimension is bigger than 32. Moreover, if we are interested in small dimensions, we can obtain a better result by choosing ε in the last inequality in a different way : in fact we have chosen $\varepsilon = (2n)^{-\frac{1}{3}}$ (replacing d_X with $\sqrt{n}$) in the asymptotic regime, otherwise one just needs to optimize ε so that it satisfies $\frac{\varepsilon}{\sqrt{(1-\varepsilon)^2 n}} = \frac{1}{n\sqrt{1-(1-\varepsilon)^2}}$; then our result becomes nontrivial when the dimension is larger than 16. In [84], Taschuk has also obtained an estimate for the Banach-Mazur distance to the cube of "small"-dimensional spaces. Precisely, he proved the following

$$R_\infty^n \leqslant \sqrt{n^2 - 2n + 2 + \frac{2}{\sqrt{n+2} - 1}}$$

We can check that our result improves on that whenever the dimension is larger than 22. We give a table of comparison of the estimates obtained from the two results. Our values are obtained as follow : write $k = (1-\varepsilon)^2 n$ so that $\varepsilon = 1 - \sqrt{\frac{k}{n}}$ then solve the equation

$$\frac{1 - \sqrt{\frac{k}{n}}}{\sqrt{k}} = \frac{1}{\sqrt{n(n-k)}}$$

Take k_0 the integer part of the solution obtained, then the upper bound will be given by

$$\frac{\sqrt{2}}{\min\left(\frac{1-\sqrt{\frac{k_0}{n}}}{\sqrt{k_0}}, \frac{1}{\sqrt{n(n-k_0)}}\right)}$$

Dimension n	Our upper bound for R_∞^n	Taschuk's upper bound for R_∞^n
200	138,564	199,002
150	108,166	149,004
100	76,158	99,006
70	56,745	69,009
40	34,641	39,017
22	20,978	21,036
21	20,494	20,039

Chapitre 4

Covariance of random matrices

4.1 Introduction

In recent years, interest in matrix valued random variables gained momentum. Many of the results dealing with real random variables and random vectors were extended to cover random matrices. Concentration inequalities like Bernstein, Hoeffding and others were obtained in the non-commutative setting ([6],[88],[54]). The methods used were mostly combination of methods from the real/vector case and some matrix inequalities like the Golden-Thompson inequality (see [15]).

Estimating the covariance matrix of a random vector has gained a lot of interest recently. Given a random vector X in $\mathbb{R}^n$, the question is to estimate $\Sigma = \mathbb{E}XX^t$. A natural way to do this, is to take $X_1, .., X_N$ independent copies of X and try to approximate Σ with the sample covariance matrix $\Sigma_N = \frac{1}{N}\sum_i X_i X_i^t$. The challenging problem is to find the minimal number of samples needed to estimate Σ. It is known using a result of Rudelson (see [67]) that for general distributions supported on the sphere of radius $\sqrt{n}$, it suffices to take $cn\log(n)$ samples. But for many distributions, a number proportional to n is sufficient. Using standard arguments, one can verify this for gaussian vectors. It was conjectured by Kannan- Lovasz- Simonovits [42] that the same result holds for log-concave distributions. This problem was solved by Adamczak et al ([4], [5]). Recently, Srivatava-Vershynin proved in [79] covariance estimate with a number of samples proportional to n, for a larger class of distributions covering the log-concave case. The method used was different from previous work in this field and the main idea was to randomize the sparsification theorem of Batson-Spielman-Srivastava [13].

Our aim in this chapter is to adapt the work of Srivastava-Vershynin to the matrix setting, replacing the vector X in the problem of the covariance matrix by an $n\times m$ random matrix A and trying to estimate $\mathbb{E}AA^t$ by the same techniques. This will be possible since in the deterministic setting, the sparsification theorem of Batson-Spielman-Srivastava [13] has been extended to a matrix setting by De Carli Silva-Harvey-Sato [26] who precisely proved the following :

Theorem 4.1. *Let $B_1, \dots, B_m$ be positive semidefinite matrices of size $n \times n$ and arbitrary rank. Set $B := \sum_i B_i$. For any $\varepsilon \in (0,1)$, there is a deterministic algorithm to construct a vector $y \in \mathbb{R}^m$ with $O(n/\varepsilon^2)$ nonzero entries such that $y \geq 0$ and*

$$B \ \preceq \ \sum_i y_i B_i \ \preceq \ (1+\varepsilon)B.$$

For an $n \times n$ matrix A, denote by $\|A\|$ the operator norm of A seen as an operator on l_2^n. The main idea is to randomize the previous result using the techniques of Srivastava-Vershynin [79]. Our problem can be formulated as follows :

Take B a positive semidefinite random matrix of size $n \times n$. How many independent copies of B are needed to approximate $\mathbb{E}B$ i.e taking $B_1, .., B_N$ independent copies of B, what is the minimal number of samples needed to make $\left\| \frac{1}{N}\sum_i B_i - \mathbb{E}B \right\|$ very small.

One can view this as an analogue to the covariance estimate of a random vector by taking for B the matrix AA^t where A is an $n \times m$ random matrix. With some regularity, we will be able to take N proportional to n. However, in the general case this is no longer true. In fact, take B uniformly distributed on $\{ne_ie_i^t\}_{i\leqslant n}$, where $(e_i)_{i\leqslant n}$ denotes the canonical basis of $\mathbb{R}^n$. It is easy to verify that $\mathbb{E}B = I_n$ and $\frac{1}{N}\sum_i B_i$ is a diagonal matrix and its diagonal coefficients are distributed as

$$\frac{n}{N}(p_1, .., p_n),$$

where p_i denotes the number of times $e_ie_i^t$ is chosen. This problem is well- studied and it is known (see [45]) that we must take $N \geqslant cn\log(n)$. This example is essentially due to Aubrun [8]. More generally, if B is a positive semidefinite matrix such that $\mathbb{E}B = I_n$ and $\|B\| \leqslant n$ almost surely, then by Rudelson's inequality in the non-commutative setting (see [57]) it is sufficient to take $cn\log(n)$ samples.

The method will work properly for a class of matrices satisfying a matrix strong regularity assumption which we denote by (MSR) and can be viewed as an analog to the property (SR)

defined in [79].

Definition 4.2. *[Property (MSR)]*
Let B be an $n \times n$ positive semidefinite random matrix such that $\mathbb{E}B = I_n$. We will say that B satisfies (MSR) if for some $c, \eta > 0$ we have :

$$\mathbb{P}(\|PBP\| \geqslant t) \leqslant \frac{c}{t^{1+\eta}} \quad \forall t \geqslant c.rank(P) \text{ and } \forall P \text{ orthogonal projection of } \mathbb{R}^n.$$

The main result of this chapter is the following :

Theorem 4.3. *Let B be an $n \times n$ positive semidefinite random matrix verifying $\mathbb{E}B = I_n$ and (MSR) for some $\eta > 0$. Then for every $\varepsilon \in (0,1)$, taking $N = C_1(\eta)$[1] $\frac{n}{\varepsilon^{2+\frac{2}{\eta}}}$ we have*

$$\mathbb{E}\left\| \frac{1}{N}\sum_{i=1}^{N} B_i - I_n \right\| \leqslant \varepsilon, \qquad \text{where } B_1, .., B_N \text{ are independent copies of } B.$$

4.2 Property (MSR) and examples

A random vector X in $\mathbb{R}^l$ is called isotropic if its covariance matrix is the identity i.e $\mathbb{E}XX^t = Id$. In [79], an isotropic random vector X in $\mathbb{R}^l$ was said to satisfy (SR) if for some $c, \eta > 0$,

$$\mathbb{P}\left(\|PX\|_2^2 \geqslant t\right) \leqslant \frac{c}{t^{1+\eta}}, \quad \forall t \geqslant c.rank(P) \text{ and } \forall P \text{ orthogonal projection of } \mathbb{R}^l.$$

Since $\|PXX^tP\| = \|PX\|_2^2$, then clearly $B = XX^t$ satisfies (MSR) if and only if X satisfies (SR). Therefore if X verifies the property (SR), applying Theorem 4.3 to $B = XX^t$, we recover the covariance estimate as stated in [79].

Let us note that (MSR) implies moment assumptions on the quadratic forms $\langle Bx, x\rangle$. To see this, first note that if $x \in \mathbb{S}^{n-1}$ then $\langle Bx, x\rangle = \|P_x B P_x\|$, where P_x is the orthogonal

1. $C_1(\eta) = (64c)^{1+\frac{2}{\eta}}(1+\frac{1}{\eta})^{\frac{2}{\eta}} \vee 64(4c)^{\frac{1}{\eta}}(32+\frac{32}{\eta})^{1+\frac{3}{\eta}} \vee 256(2c)^{\frac{3}{2}+\frac{2}{\eta}}(16+\frac{16}{\eta})^{\frac{4}{\eta}}$

projection on span(x). Now, by integration of tails we have for $1 < q < 1+\eta$,

$$\begin{aligned}\mathbb{E}\langle Bx,x\rangle^q &= \int_0^\infty \mathbb{P}(\langle Bx,x\rangle > t).qt^{q-1}dt\\ &= \int_0^c \mathbb{P}(\langle Bx,x\rangle > t).qt^{q-1}dt\\ &\quad + \int_c^\infty \mathbb{P}(\langle Bx,x\rangle > t).qt^{q-1}dt\\ &\leqslant \int_0^c qt^{q-1}dt + \int_c^\infty \frac{c}{t^{1+\eta}}.qt^{q-1}dt\\ &\leqslant c(q,\eta)\end{aligned}$$

Moreover, property (MSR) implies regularity assumption on the eigenvalues of the matrix B. Indeed, for any orthogonal projection of rank k one can write

$$\|PBP\| \geqslant \min_{\substack{F\subset\mathbb{R}^n\\ \dim F=k}} \max_{x\in F}\langle Bx,x\rangle = \lambda_{n-k+1}(B),$$

where the last equality is given by the Courant-Fisher minimax formula (see [25]). Therefore, property (MSR) implies the following : for some $c,\eta>0$,

$$\mathbb{P}\left(\lambda_{n-k+1}(B)\geqslant t\right) \leqslant \frac{c}{t^{1+\eta}}, \quad \forall t\geqslant c.k \text{ and } \forall k\leqslant n.$$

We may now discuss some examples for applications of the main result. Let us first replace (MSR) with a stronger, but easier to manipulate, property which we denote by $(MSR*)$. If B is an $n\times n$ positive semidefinite random matrix such that $\mathbb{E}B = Id$, we will say that B satisfies $(MSR*)$ if for some $c,\eta>0$:

$$\mathbb{P}(\mathrm{Tr}(PB)\geqslant t) \leqslant \frac{c}{t^{1+\eta}} \quad \forall t\geqslant c.rank(P) \text{ and } \forall P \text{ orthogonal projection of } \mathbb{R}^n.$$

Note that since $\|PBP\| \leqslant \mathrm{Tr}\,(PBP) = \mathrm{Tr}\,(PB)$, then $(MSR*)$ is clearly stronger than (MSR).

4.2.1 $(2+\varepsilon)$-moments for the spectrum

As we mentioned before, (MSR), one can see that it implies regularity assumptions on the eigenvalues of B. Putting some independence in the spectral decomposition of B, we will only need to use the regularity of the eigenvalues. To be more precise, we have the following :

Proposition 4.4. *Let $B = UDU^*$ be the spectral decomposition of an $n\times n$ symmetric positive* *semidefinite random matrix. Denote $(\alpha_j)_{j\leqslant n}$ the diagonal entries of D. Suppose that U and D are independent and that $(\alpha_j)_{j\leqslant n}$ are independent and satisfy the following :*

$$\forall i \leqslant n,\ \mathbb{E}\alpha_i = 1 \quad \text{and} \quad (\mathbb{E}\alpha_i^p)^{\frac{1}{p}} \leqslant c,$$

for some $p > 2$. Then B satisfies $(MSR*)$.

Proof. First note that since U and D are independent and $\mathbb{E}\alpha_i = 1$, then $\mathbb{E}B = Id$. Let $k > 0$ and P be an orthogonal projection of rank k on $\mathbb{R}^n$, then $Q = U^*PU$ is a random orthogonal projection of rank k independent of D. Note that $\mathrm{Tr}\,(PB) = \sum_{i\leqslant n} q_{ii}\alpha_i$, and now using Markov's inequality we have for $t > k$,

$$\begin{aligned}
\mathbb{P}\left\{\mathrm{Tr}\,(PB) \geqslant t\right\} &= \mathbb{P}\left\{\sum_{i\leqslant n} q_{ii}(\alpha_i - 1) \geqslant t - k\right\} \\
&\leqslant \mathbb{P}\left\{\left|\sum_{i\leqslant n} q_{ii}(\alpha_i - 1)\right| \geqslant t - k\right\} \\
&\leqslant \frac{1}{(t-k)^p}\mathbb{E}\left|\sum_{i\leqslant n} q_{ii}(\alpha_i - 1)\right|^p
\end{aligned}$$

Looking at the expectation with respect to D and using Rosenthal's inequality (see [63]) we get

$$\mathbb{E}_D\left|\sum_{i\leqslant n} q_{ii}(\alpha_i - 1)\right|^p \leqslant C(p)\max\left\{\sum_{i\leqslant n} q_{ii}^p\mathbb{E}|\alpha_i - 1|^p, \left(\sum_{i\leqslant n} q_{ii}^2\mathbb{E}|\alpha_i - 1|^2\right)^{\frac{p}{2}}\right\}$$

Taking in account that $q_{ii} \leqslant 1$, which implies that for any $l \geqslant 1$, $\sum_i q_{ii}^l \leqslant k$, we deduce that

$$\mathbb{E}\left|\sum_{i\leqslant n} q_{ii}(\alpha_i - 1)\right|^p \leqslant C(p)k^{\frac{p}{2}}$$

Instead of Rosenthal's inequality, we could have used a symmetrization argument alongside Khintchine's inequality to get the estimate above.
One can easily conclude that B satisfies $(MSR*)$ with $\eta = \frac{p}{2} - 1$. □

Applying Theorem 4.3, we can deduce the following proposition :

Proposition 4.5. *Let $B = UDU^*$ be the spectral decomposition of an $n\times n$ symmetric positive semidefinite random matrix. Denote $(\alpha_j)_{j\leqslant n}$ the diagonal entries of D. Suppose that U and D*

are independent and that $(\alpha_j)_{j\leqslant n}$ are independent and satisfy the following :

$$\forall i \leqslant n,\ \mathbb{E}\alpha_i = 1 \quad and \quad (\mathbb{E}\alpha_i^p)^{\frac{1}{p}} \leqslant c,$$

for some $p > 2$. Let $\varepsilon \in (0,1)$, then taking $N = C(p)\frac{n}{\varepsilon^{\frac{2p}{p-2}}}$ we have

$$\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N} B_i - I_n\right\| \leqslant \varepsilon, \quad where\ B_1, .., B_N\ are\ independent\ copies\ of\ B.$$

4.2.2 From (SR) to (MSR)

We will show how to jump from property (SR) dealing with vectors to the property $(MSR*)$ dealing with matrices.

Proposition 4.6. *Let A be an $n\times m$ random matrix and denote by $(C_i)_{i\leqslant m}$ its columns. Suppose that $A'^t = \sqrt{m}(C_1^t, .., C_m^t)$ is an isotropic random vector in $\mathbb{R}^{nm}$ which satisfies property (SR). Then $B = AA^t$ verifies $\mathbb{E}B = I_n$ and Property $(MSR*)$.*

Proof. For $l \leqslant nm$, one can write $l = (j-1)n + i$ with $1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant m$, so that the coordinates of A' are given by $a'_l = \sqrt{m}a_{i,j}$, and since A' is isotropic we get $\mathbb{E}a_{i,j}a_{r,s} = \frac{1}{m}\delta_{(i,j),(r,s)}$. The terms of B are given by $b_{i,j} = \sum_{s=1}^{m} a_{i,s}a_{j,s}$. We deduce that $\mathbb{E}b_{i,j} = \delta_{i,j}$ and therefore $\mathbb{E}B = I_n$. Let P be an orthogonal projection of $\mathbb{R}^n$ and put $P' = I_m \otimes P$ i.e P' is an $nm \times nm$ matrix of the form

$$P' = \begin{pmatrix} P & 0 & \dots & 0 \\ 0 & P & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & P \end{pmatrix}$$

Clearly we have $\|P'A'\|_2^2 = m\mathrm{Tr}(PB)$ and $rank(P') = m.rank(P)$.
Let $t \geqslant c.rank(P)$ then $mt \geqslant c.rank(P')$ and by property (SR) we have :

$$\mathbb{P}\left(\|P'A'\|_2^2 \geqslant mt\right) \leqslant \frac{c}{(mt)^{1+\eta}}$$

This means that

$$\mathbb{P}\left(\mathrm{Tr}(PB) \geqslant t\right) \leqslant \frac{c}{(mt)^{1+\eta}}$$

and therefore B satisfies $(MSR*)$. □

4.2.3 Bounded matrices

Let A be an $n \times n$ symmetric centered random matrix satisfying $\|A\| \leqslant M$ almost surely. Set $B = \frac{1}{M}A + I_n$, then B is symmetric positive semidefinite and $\mathbb{E}B = I_n$.
Since $\|PBP\| \leqslant 2$ a.s, then B satisfies (MSR) almost surely. Applying Theorem 4.3 to B, we have for $N \geqslant c\frac{n}{\varepsilon^2}$, that $\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N} B_i - I_n\right\|_2 \leqslant \varepsilon$ which means

$$\mathbb{E}\lambda_{max}\left(\frac{1}{N}\sum_{i=1}^{N} A_i\right) \leqslant \varepsilon M \quad \textit{and} \quad \mathbb{E}\lambda_{min}\left(\frac{1}{N}\sum_{i=1}^{N} A_i\right) \geqslant -\varepsilon M$$

This is however weaker than matrix Bernstein's inequality, see [88]. Although the conclusion is not trivial, but it is not satisfactory since we do not know how to use the fact that A is bounded almost surely in order to reduce the number N of copies needed.

4.3 Proof of Theorem 4.3

We first introduce a regularity assumption on the moments which we denote by (MWR) :

$$\exists p > 1 \;\text{ such that }\; \mathbb{E}\left\langle Bx, x\right\rangle^p \leqslant C_p \quad \forall x \in S^{n-1}.$$

Note that by a simple integration of tails, (MSR) (with P a rank one projection) implies (MWR) with $p < 1 + \eta$.

The proof of Theorem 4.3 is based on two theorems dealing with the smallest and largest eigenvalues of $\frac{1}{N}\sum_{i=1}^{N} B_i$.

Theorem 4.7. *Let B_i be $n \times n$ independent positive semidefinite random matrices verifying $\mathbb{E}B_i = I_n$ and (MWR) .*
Let $\varepsilon \in (0,1)$, then for

$$N \geqslant 16\left(16C_p\right)^{\frac{1}{p-1}} \frac{n}{\varepsilon^{\frac{2p-1}{p-1}}}$$

we get

$$\mathbb{E}\lambda_{min}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \geqslant 1 - \varepsilon$$

Remark 4.8. *The proof yields a more general estimate ; precisely if $h = \frac{n}{N}$ then*

$$\mathbb{E}\lambda_{min}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \geqslant 1 - c(p)\max\left\{h^{\frac{p-1}{2p-1}}, h\right\}$$

Theorem 4.9. *Let B_i be $n \times n$ independent positive semidefinite random matrices verifying $\mathbb{E}B_i = I_n$ and (MSR).*
For $\varepsilon \in (0,1)$ and $N \geqslant C_2(\eta)$ [2] $\frac{n}{\varepsilon^{2+\frac{2}{\eta}}}$ we have

$$\mathbb{E}\lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \leqslant 1+\varepsilon$$

Remark 4.10. *The proof yields a more general estimate ; Precisely if $h = \frac{n}{N}$ then*

$$\mathbb{E}\lambda_{max}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \leqslant 1 + c(\eta)\max\left\{h^{\frac{\eta}{2+2\eta}}, h\right\}$$

Combining this with the previous remark, for any $B_1, ..., B_N$ $n \times n$ independent positive semi-definite random matrices verifying $\mathbb{E}B_i = I_n$ and (MSR), we have

$$1 - c(\eta)\max\left\{h^{\frac{\eta}{2+2\eta}}, h\right\} \leqslant \mathbb{E}\lambda_{min}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \leqslant \mathbb{E}\lambda_{max}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \leqslant 1 + c(\eta)\max\left\{h^{\frac{\eta}{2+2\eta}}, h\right\}$$

We will give the proof of these two theorems in sections 4.4 and 4.5 respectively. We need also a simple lemma :

Lemma 4.11. *Let $1 < r \leqslant 2$ and $Z_1, ..., Z_N$ be independent positive random variables with $\mathbb{E}Z_i = 1$ and satisfying $(\mathbb{E}Z_i^r)^{\frac{1}{r}} \leqslant M$ Then*

$$\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^{N} Z_i - 1\right| \leqslant \frac{2M}{N^{\frac{r-1}{r}}}.$$

Proof. Let $(\varepsilon_i)_{i\leqslant N}$ independent ± 1 Bernoulli variables. By symmetrization and Jensen's inequality we can write

$$\begin{aligned}\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^{N} Z_i - 1\right| &\leqslant \frac{2}{N}\mathbb{E}\left|\sum_{i=1}^{N}\varepsilon_i Z_i\right| \leqslant \frac{2}{N}\mathbb{E}\left(\sum_{i=1}^{N} Z_i^2\right)^{\frac{1}{2}} \\ &\leqslant \frac{2}{N}\mathbb{E}\left(\sum_{i=1}^{N} Z_i^r\right)^{\frac{1}{r}} \leqslant \frac{2}{N}\left(\sum_{i=1}^{N}\mathbb{E}Z_i^r\right)^{\frac{1}{r}} \leqslant \frac{2M}{N^{\frac{r-1}{r}}}\end{aligned}$$

□

Proof of Theorem 4.3. Take $N \geqslant c(\eta)\frac{n}{\varepsilon^{2+\frac{2}{\eta}}}$ satisfying conditions of Theorem 4.7 (with $p = 1+\frac{\eta}{2}$) and Theorem 4.9. Note that by the triangle inequality

2. $C_2(\eta) = 16c^{\frac{1}{\eta}}(32+\frac{32}{\eta})^{1+\frac{3}{\eta}} \vee 16\sqrt{2}(4c)^{\frac{3}{2}+\frac{2}{\eta}}(8+\frac{8}{\eta})^{\frac{4}{\eta}}$

$$\left\| \frac{1}{N}\sum_{i=1}^N B_i - I_n \right\| \leqslant \left\| \frac{1}{N}\sum_{i=1}^N B_i - \frac{1}{n}\mathrm{Tr}\left(\frac{1}{N}\sum_{i=1}^N B_i\right) I_n \right\| + \left\| \frac{1}{n}\mathrm{Tr}\left(\frac{1}{N}\sum_{i=1}^N B_i\right) I_n - I_n \right\|$$

$$:= \alpha + \beta$$

Observe that

$$\begin{aligned}\alpha &= \max\left|\lambda\left(\frac{1}{N}\sum_{i=1}^N B_i - \frac{1}{n}\mathrm{Tr}\left(\frac{1}{N}\sum_{i=1}^N B_i\right) I_n\right)\right| \\ &= \max\left[\lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^N B_i\right) - \frac{1}{n}\mathrm{Tr}\left(\frac{1}{N}\sum_{i=1}^N B_i\right), \frac{1}{n}\mathrm{Tr}\left(\frac{1}{N}\sum_{i=1}^N B_i\right) - \lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^N B_i\right)\right]\end{aligned}$$

Since the two terms in the max are non-negative, then one can bound the max by the sum of the two terms. More precisely, we get $\alpha \leqslant \lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^N B_i\right) - \lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^N B_i\right)$ and by Theorem 4.7 and Theorem 4.9 we deduce that $\mathbb{E}\alpha \leqslant 2\varepsilon$.

Note that

$$\beta = \left|\frac{1}{N}\sum_{i=1}^N \frac{\mathrm{Tr}(B_i)}{n} - 1\right| = \left|\frac{1}{N}\sum_{i=1}^N Z_i - 1\right|,$$

where $Z_i = \frac{\mathrm{Tr}(B_i)}{n}$. Since B_i satisfies (MWR), then taking $r = \min(2, 1+\frac{\eta}{2})$ we have

$$\forall i \leqslant N, \quad (\mathbb{E}Z_i^r)^{\frac{1}{r}} \leqslant \frac{1}{n}\sum_{j=1}^n (\mathbb{E}\langle B_i e_j, e_j\rangle^r)^{\frac{1}{r}} \leqslant c(\eta).$$

Therefore Z_i satisfies the conditions of Lemma 4.11 and we deduce that $\mathbb{E}\beta \leqslant \varepsilon$ by the choice of N.

As a conclusion

$$\mathbb{E}\left\| \frac{1}{N}\sum_{i=1}^N B_i - I_n \right\| \leqslant \mathbb{E}\alpha + \mathbb{E}\beta \leqslant 3\varepsilon$$

□

4.4 Proof of Theorem 4.7

Given A an $n \times n$ positive semidefinite matrix such that all its eigenvalues are greater than a lower barrier $l_A = l$ i.e $A \succ l.I_n$, define the corresponding potential function to be $\phi_l(A) = \mathrm{Tr}(A - l.I_n)^{-1}$.
The proof of Theorem 4.7 is based on the following result which will be proved in section 4.6 :

Theorem 4.12. *Let $A \succ l.I_n$ and $\phi_l(A) \leqslant \phi$, B a positive semidefinite random matrix satisfying $\mathbb{E}B = I_n$ and Property (MWR) with some $p > 1$.*
Let $\varepsilon \in (0,1)$, if

$$\phi \leqslant \frac{1}{4\,(8C_p)^{\frac{1}{p-1}}}\varepsilon^{\frac{p}{p-1}}$$

then there exists l' a random variable such that

$$A + B \succ l'.I_n, \quad \phi_{l'}(A+B) \leqslant \phi_l(A) \quad \text{and} \quad \mathbb{E}l' \geqslant l + 1 - \varepsilon.$$

Proof of Theorem 4.7. We start with $A_0 = 0$ and $l_0 = -\frac{n}{\phi}$ so that $\phi_{l_0}(A_0) = -\frac{n}{l_0} = \phi$.
Applying Theorem 4.12, one can find l_1 such that

$$A_1 = A_0 + B_1 \succ l_1.I_n, \quad \phi_{l_1}(A_1) \leqslant \phi_{l_0}(A_0)$$

and

$$\mathbb{E}l_1 \geqslant l_0 + 1 - \varepsilon$$

Now apply Theorem 4.12 once again to find l_2 such that

$$A_2 = A_1 + B_2 \succ l_2.I_n, \quad \phi_{l_2}(A_2) \leqslant \phi_{l_1}(A_1)$$

and

$$\mathbb{E}l_2 \geqslant l_1 + 1 - \varepsilon \geqslant l_0 + 2(1-\varepsilon)$$

After N steps, we get $\mathbb{E}\lambda_{\min}(A_N) \geqslant \mathbb{E}l_N \geqslant l_0 + N(1-\varepsilon)$. Therefore,

$$\mathbb{E}\lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \geqslant 1 - \varepsilon - \frac{n}{N\phi}$$

Taking $N = \frac{n}{\varepsilon\phi}$, we get $\mathbb{E}\lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \geqslant 1 - 2\varepsilon$.

□

4.5 Proof of Theorem 4.9

Given A an $n \times n$ positive semidefinite matrix such that all eigenvalues of A are less than an upper barrier $u_A = u$ i.e $A \prec u.I_n$, define the corresponding potential function to be $\psi_u(A) = \mathrm{Tr}\,(u.I_n - A)^{-1}$.
The proof of Theorem 4.9 is based on the following result which will be proved in section 4.7 :

Theorem 4.13. *Let $A \prec u.I_n$ and $\psi_u(A) \leqslant \psi$, B a positive semidefinite random matrix satisfying $\mathbb{E}B = I_n$ and Property (MSR).*
Let $\varepsilon \in (0,1)$, if

$$\psi \leqslant C_3(\eta)\,^3\varepsilon^{1+\frac{2}{\eta}}$$

then there exists u' a random variable such that

$$A + B \prec u'.I_n, \quad \psi_{u'}(A+B) \leqslant \psi_u(A) \quad \textit{and} \quad \mathbb{E}u' \leqslant u + 1 + \varepsilon.$$

Proof of Theorem 4.9. We start with $A_0 = 0$, $u_0 = \frac{n}{\psi}$ so that $\psi_{u_0}(A_0) = \psi$.
Applying Theorem 4.13, one can find u_1 such that

$$A_1 = A_0 + B_1 \prec u_1.I_n, \quad \psi_{u_1}(A_1) \leqslant \psi_{u_0}(A_0) \;\text{ and }\; \mathbb{E}u_1 \leqslant u_0 + 1 + \varepsilon.$$

Now apply Theorem 4.13 once again to find u_2 such that

$$A_2 = A_1 + B_2 \prec u_2.I_n, \quad \psi_{u_2}(A_2) \leqslant \psi_{u_1}(A_1) \;\text{ and }\; \mathbb{E}u_2 \leqslant u_1 + 1 + \varepsilon.$$

After N steps we get $\mathbb{E}\lambda_{\max}\left(\sum_{i=1}^{N} B_i\right) \leqslant \mathbb{E}u_N \leqslant u_0 + N(1+\varepsilon)$.

3. $C_3(\eta) = \left[8(2c)^{\frac{1}{\eta}}(16 + \frac{16}{\eta})^{1+\frac{3}{\eta}}\right]^{-1} \wedge \left[16(2c)^{\frac{3}{2}+\frac{2}{\eta}}(8 + \frac{8}{\eta})^{\frac{4}{\eta}}\right]^{-1}$

Taking $N \geqslant \frac{n}{\varepsilon\psi} = c(\eta)\frac{n}{\varepsilon^{2+\frac{2}{\eta}}}$, we deduce that

$$\mathbb{E}\lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \leqslant 1 + 2\varepsilon$$

□

4.6 Proof of Theorem 4.12

4.6.1 Notations

We are looking for a random variable l' of the form $l + \delta$ where δ is a positive random variable playing the role of the shift.

If in addition $A \succ (l+\delta).I_n$, we will note :
$L_\delta = A - (l+\delta).I_n \succ 0$ so that $\mathrm{Tr}\left(B^{\frac{1}{2}}(A - (l+\delta).I_n)^{-1}B^{\frac{1}{2}}\right) = \left\langle L_\delta^{-1}, B\right\rangle$.

$\lambda_1, .., \lambda_n$ will denote the eigenvalues of A and $v_1, .., v_n$ the corresponding eigenvectors. Note that $(v_i)_{i\leqslant n}$ are also the eigenvectors of L_δ^{-1} corresponding to the eigenvalues $\frac{1}{\lambda_i-(l+\delta)}$.

4.6.2 Finding the shift

To find sufficient conditions for such δ exists, we need a matrix extension of Lemma 3.4 in [13] which, up to a minor change, is essentially contained in Lemma 20 in [26] and we formulate it here in Lemma 4.15. The method uses the Sherman-Morrison-Woodbury formula :

Lemma 4.14. *Let E be an $n \times n$ invertible matrix, C a $k \times k$ invertible matrix, U an $n \times k$ matrix and V a $k \times n$ matrix. Then we have :*

$$(E + UCV)^{-1} = E^{-1} - E^{-1}U(C^{-1} + VE^{-1}U)^{-1}VE^{-1}$$

Proof. Write

$$\begin{aligned}
&(E+UCV)\cdot\left[E^{-1}-E^{-1}U(C^{-1}+VE^{-1}U)^{-1}VE^{-1}\right]\\
&=Id+UCVE^{-1}-(U+UCVE^{-1}U)(C^{-1}+VE^{-1}U)^{-1}VE^{-1}\\
&=Id+UCVE^{-1}-UC(C^{-1}+VE^{-1}U)(C^{-1}+VE^{-1}U)^{-1}VE^{-1}\\
&=Id+UCVE^{-1}-UCVE^{-1}=Id
\end{aligned}$$

In a similar way, $[E^{-1}-E^{-1}U(C^{-1}+VE^{-1}U)^{-1}VE^{-1}]\cdot(E+UCV)=Id$ and this finishes the proof. □

Lemma 4.15. *Let A as above satisfying $A\succ l.I_n$. Suppose that one can find $\delta>0$ verifying $\delta<\frac{1}{\|L_0^{-1}\|}$ and*

$$\frac{\left\langle L_\delta^{-2},B\right\rangle}{\phi_{l+\delta}(A)-\phi_l(A)}-\left\|B^{\frac12}L_\delta^{-1}B^{\frac12}\right\|\geqslant 1$$

Then

$$\lambda_{\min}(A+B)>l+\delta\quad\text{and}\quad\phi_{l+\delta}(A+B)\leqslant\phi_l(A).$$

Proof. First note that $\frac{1}{\|L_0^{-1}\|}=\lambda_{\min}(A)-l$, so the first condition on δ implies that $\lambda_{\min}(A)>l+\delta$. Now using Sherman-Morrison-Woodbury formula with $E=L_\delta,\ U=V=B^{\frac12},\ C=I_n$ we get

$$\begin{aligned}
\phi_{l+\delta}(A+B)&=\mathrm{Tr}\,(L_\delta+B)^{-1}\\
&=\phi_{l+\delta}(A)-\mathrm{Tr}\left(L_\delta^{-1}B^{\frac12}\left(I_n+B^{\frac12}L_\delta^{-1}B^{\frac12}\right)^{-1}B^{\frac12}L_\delta^{-1}\right)\\
&\leqslant\phi_{l+\delta}(A)-\frac{\left\langle L_\delta^{-2},B\right\rangle}{1+\left\|B^{\frac12}L_\delta^{-1}B^{\frac12}\right\|}
\end{aligned}$$

Rearranging the hypothesis, we get $\phi_{l+\delta}(A+B)\leqslant\phi_l(A)$.

□

Since $\|L_0^{-1}\|\leqslant\mathrm{Tr}\left(L_0^{-1}\right)=\phi_l(A)$ and $\left\|B^{\frac12}L_\delta^{-1}B^{\frac12}\right\|\leqslant\left\langle L_\delta^{-1},B\right\rangle$ then in order to satisfy conditions of Lemma 4.15, we may search for δ satisfying :

$$\delta<\frac{1}{\phi_l(A)}\quad\text{and}\quad\frac{\left\langle L_\delta^{-2},B\right\rangle}{\phi_{l+\delta}(A)-\phi_l(A)}-\left\langle L_\delta^{-1},B\right\rangle\geqslant 1\tag{4.1}$$

For $t<\frac{1}{\phi}$, let us note :

$$q_1(t,B) = \left\langle L_t^{-1}, B\right\rangle = \mathrm{Tr}\left(B(A-(l+t).I_n)^{-1}\right)$$

and

$$q_2(t,B) = \frac{\left\langle L_t^{-2}, B\right\rangle}{\mathrm{Tr}(L_t^{-2})} = \frac{\mathrm{Tr}\,(B(A-(l+t).I_n)^{-2})}{\mathrm{Tr}\,(A-(l+t).I_n)^{-2}}$$

We have already seen in Lemma 4.15 that if $t < \frac{1}{\phi} \leqslant \frac{1}{\|L_0^{-1}\|}$ then $A \succ (l+t).I_n$ so the definitions above make sense. Since we have

$$\begin{aligned}\phi_{l+\delta}(A) - \phi_l(A) &= \mathrm{Tr}(A-(l+\delta).I_n)^{-1} - \mathrm{Tr}(A-l.I_n)^{-1}\\ &= \delta\mathrm{Tr}((A-(l+\delta).I_n)^{-1}(A-l.I_n)^{-1})\\ &\leqslant \delta\mathrm{Tr}(A-(l+\delta).I_n)^{-2},\end{aligned}$$

then in order to have (4.1), it will be sufficient to choose δ satisfying $\delta < \frac{1}{\phi}$ and

$$\frac{1}{\delta}q_2(\delta,B) - q_1(\delta,B) \geqslant 1 \tag{4.2}$$

Note that q_1 and q_2 can be expressed as follows

$$q_1(t,B) = \sum_{i=1}^{n}\frac{\langle Bv_i, v_i\rangle}{\lambda_i - l - t} \quad \text{and} \quad q_2(t,B) = \frac{\sum_i \frac{\langle Bv_i, v_i\rangle}{(\lambda_i - l - t)^2}}{\sum_i(\lambda_i - l - t)^{-2}}$$

Since $\phi_l(A) = \sum_{i=1}^{n}(\lambda_i - l)^{-1} \leqslant \phi$ then $(\lambda_i - l).\phi \geqslant 1$ for all i, and we have

$$(1-t.\phi)(\lambda_i - l) = \lambda_i - l - t.(\lambda_i - l).\phi \leqslant \lambda_i - l - t \leqslant \lambda_i - l$$

therefore

$$q_1(t,B) \leqslant (1-t.\phi)^{-1}q_1(0,B)$$

and

$$(1-t.\phi)^2 q_2(0,B) \leqslant q_2(t,B) \leqslant (1-t.\phi)^{-2}q_2(0,B)$$

Lemma 4.16. *Let $s \in (0,1)$ and take $\delta = (1-s)^3 q_2(0,B)\mathbf{1}_{\{q_1(0,B)\leqslant s\}}\mathbf{1}_{\{q_2(0,B)\leqslant \frac{s}{\phi}\}}$. Then $A+B \succ (l+\delta).I_n$ and $\phi_{l+\delta}(A+B) \leqslant \phi_l(A)$.*

Proof. As stated before, it is sufficient to check that $\delta < \frac{1}{\phi}$ and $\frac{1}{\delta}q_2(\delta,B) - q_1(\delta,B) \geqslant 1$.
If $q_1(0,B) \geqslant s$ or $q_2(0,B) \geqslant \frac{s}{\phi}$ then $\delta = 0$ and there is nothing to prove since $\phi_l(A+B) \leqslant \phi_l(A)$.

In the other case i.e when $q_1(0,B) \leqslant s$ and $q_2(0,B) \leqslant \frac{s}{\phi}$, we have $\delta = (1-s)^3 q_2(0,B)$. Therefore, $\delta \leqslant (1-s)^3 \frac{s}{\phi} < \frac{1}{\phi}$ and

$$\begin{aligned} \frac{1}{\delta} q_2(\delta,B) - q_1(\delta,B) &= \frac{1}{(1-s)^3 q_2(0,B)} q_2(\delta,B) - q_1(\delta,B) \\ &\geqslant \frac{1}{(1-s)^3 q_2(0,B)} (1-\delta\phi)^2 q_2(0,B) - (1-\delta\phi)^{-1} q_1(0,B) \\ &\geqslant \frac{(1-s)^2}{(1-s)^3} - \frac{s}{(1-s)} = 1 \end{aligned}$$

□

4.6.3 Estimating the random shift

Now that we have found δ, we will estimate $\mathbb{E}\delta$ using the property (MWR). We will start by stating some basic facts about q_1 and q_2.

Proposition 4.17. *Let, as above, $A \succ l.I_n$ and $\phi_l(A) \leqslant \phi$, B satisfying (MWR). Then we have the following :*

1. $\mathbb{E}q_1(0,B) = \phi_l(A) \leqslant \phi$ *and* $\mathbb{E}q_1(0,B)^p \leqslant C_p \phi^p$.
2. $\mathbb{E}q_2(0,B) = 1$ *and* $\mathbb{E}q_2(0,B)^p \leqslant C_p$.
3. $\mathbb{P}(q_1(0,B) \geqslant u) \leqslant C_p(\frac{\phi}{u})^p$ *and* $\mathbb{P}(q_2(0,B) \geqslant u) \leqslant \frac{C_p}{u^p}$.

Proof. Since $\mathbb{E}B = I_n$ then $\mathbb{E}q_1(0,B) = \phi_l(A)$ and $\mathbb{E}q_2(0,B) = 1$.
Now using the triangle inequality and Property (MWR) we get

$$(\mathbb{E}q_1(0,B)^p)^{\frac{1}{p}} = \left[\mathbb{E}\left(\sum_{i=1}^n \frac{\langle Bv_i, v_i\rangle}{\lambda_i - l}\right)^p\right]^{\frac{1}{p}} \leqslant \sum_{i=1}^n \frac{(\mathbb{E}\langle Bv_i, v_i\rangle^p)^{\frac{1}{p}}}{\lambda_i - l} \leqslant \sum_{i=1}^n \frac{C_p^{\frac{1}{p}}}{\lambda_i - l} \leqslant C_p^{\frac{1}{p}} \phi$$

With the same argument we prove that $\mathbb{E}q_2(0,B)^p \leqslant C_p$. The third part of the proposition follows by Markov's inequality. □

Lemma 4.18. *If δ is as in Lemma 4.16. Then*

$$\mathbb{E}\delta \geqslant (1-s)^3 \left[1 - 2C_p \left(\frac{\phi}{s}\right)^{p-1}\right]$$

Proof. Using the above proposition and Hölder's inequality with $\frac{1}{p}+\frac{1}{q}=1$ we get :

$$\begin{aligned}
\mathbb{E}\delta &= \mathbb{E}(1-s)^3 q_2(0,B)\mathbf{1}_{\{q_1(0,B)\leqslant s\}}\mathbf{1}_{\{q_2(0,B)\leqslant \frac{s}{\phi}\}} \\
&= (1-s)^3\left[\mathbb{E}q_2(0,B) - \mathbb{E}q_2(0,B)\mathbf{1}_{\{q_1(0,B)>s \text{ or } q_2(0,B)>\frac{s}{\phi}\}}\right] \\
&\geqslant (1-s)^3\left[1-(\mathbb{E}q_2(0,B)^p)^{\frac{1}{p}}.\left(\mathbb{P}\left\{q_1(0,B)>s \text{ or } q_2(0,B)>\frac{s}{\phi}\right\}\right)^{\frac{1}{q}}\right] \\
&\geqslant (1-s)^3\left[1-C_p^{\frac{1}{p}}\left(C_p\left(\frac{\phi}{s}\right)^p + C_p\left(\frac{\phi}{s}\right)^p\right)^{\frac{1}{q}}\right] \\
&\geqslant (1-s)^3\left[1-2C_p\left(\frac{\phi}{s}\right)^{p-1}\right]
\end{aligned}$$

□

Now it remains to make good choice of s and ϕ in order to finish the proof of Theorem 4.12. Take $l' = l+\delta$, the choice of δ being as before with $s=\frac{\varepsilon}{4}$.
As we have seen, we get $A+B \succ l'.I_n$ and $\phi_{l'}(A+B)\leqslant \phi_l(A)$. Moreover,

$$\mathbb{E}l' = l + \mathbb{E}\delta \geqslant l + (1-s)^3\left[1-2C_p\left(\frac{\phi}{s}\right)^{p-1}\right] \geqslant 1-\varepsilon,$$

by the choice of ϕ. This ends the proof of Theorem 4.12.

4.7 Proof of Theorem 4.13

4.7.1 Notations

We are looking for a random variable u' of the form $u+\Delta$ where Δ is a positive random variable playing the role of the shift.

We will note $U_t = (u+t).I_n - A$ so that $\mathrm{Tr}\left(B^{\frac{1}{2}}((u+t).I_n - A)^{-1}B^{\frac{1}{2}}\right) = \left\langle {U_t}^{-1}, B\right\rangle$.

As before, $\lambda_1,..,\lambda_n$ will denote the eigenvalues of A and $v_1,..,v_n$ the corresponding eigenvectors. Note that $(v_i)_{i\leqslant n}$ are also the eigenvectors of U_t^{-1} corresponding to the eigenvalues $\frac{1}{u+t-\lambda_i}$.

4.7.2 Finding the shift

To find sufficient conditions for such Δ exists, we need a matrix extension of Lemma 3.3 in [13] which, up to a minor change, is essentially contained in Lemma 19 in [26]. For the sake of completeness, we include the proof.

Lemma 4.19. *Let A be as above satisfying $A \prec u.I_n$. Suppose that one can find $\Delta > 0$ verifying*

$$\frac{\left\langle U_\Delta^{-2}, B\right\rangle}{\psi_u(A) - \psi_{u+\Delta}(A)} + \left\| B^{\frac{1}{2}} U_\Delta^{-1} B^{\frac{1}{2}} \right\| \leqslant 1 \tag{4.3}$$

Then

$$A + B \prec (u+\Delta).I_n \quad \textit{and} \quad \psi_{u+\Delta}(A+B) \leqslant \psi_u(A).$$

Proof. Since $\left\langle U_\Delta^{-2}, B\right\rangle$ and $\psi_u(A) - \psi_{u+\Delta}(A)$ are positive, then by (4.3) we have

$$\left\| B^{\frac{1}{2}} U_\Delta^{-1} B^{\frac{1}{2}} \right\| < 1 \quad \text{and} \quad \frac{\left\langle U_\Delta^{-2}, B\right\rangle}{1 - \left\| B^{\frac{1}{2}} U_\Delta^{-1} B^{\frac{1}{2}} \right\|} \leqslant \psi_u(A) - \psi_{u+\Delta}(A)$$

First note that $\left\| B^{\frac{1}{2}} U_\Delta^{-1} B^{\frac{1}{2}} \right\| = \left\| U_\Delta^{-\frac{1}{2}} B U_\Delta^{-\frac{1}{2}} \right\| < 1$, so $U_\Delta^{-\frac{1}{2}} B U_\Delta^{-\frac{1}{2}} \prec I_n$. Therefore we get $B \prec U_\Delta$ which means that $A + B \prec (u+\Delta).I_n$.
Now using the Sherman-Morrison-Woodbury (see Lemma 4.14) with $E = U_\Delta, U = V = B^{\frac{1}{2}}, C = I_n$ we get :

$$\begin{aligned}
\psi_{u+\Delta}(A+B) &= \operatorname{Tr}\left(U_\Delta - B\right)^{-1} \\
&= \psi_{u+\Delta}(A) + \operatorname{Tr}\left(U_\Delta^{-1} B^{\frac{1}{2}} \left(I_n - B^{\frac{1}{2}} U_\Delta^{-1} B^{\frac{1}{2}} \right)^{-1} B^{\frac{1}{2}} U_\Delta^{-1} \right) \\
&\leqslant \psi_{u+\Delta}(A) + \frac{\left\langle U_\Delta^{-2}, B\right\rangle}{1 - \left\| U_\Delta^{-\frac{1}{2}} B U_\Delta^{-\frac{1}{2}} \right\|} \leqslant \psi_u(A)
\end{aligned}$$

□

We may now find Δ satisfying (4.3). Let us note :

$$Q_1(t,B) = \left\| B^{\frac{1}{2}} U_t^{-1} B^{\frac{1}{2}} \right\| = \left\| B^{\frac{1}{2}} \left((u+t).I_n - A\right)^{-1} B^{\frac{1}{2}} \right\|$$

and

$$Q_2(t,B) = \frac{\left\langle U_t^{-2}, B\right\rangle}{\psi_u(A) - \psi_{u+t}(A)} = \frac{\operatorname{Tr}\left(B\left((u+t).I_n - A\right)^{-2}\right)}{\psi_u(A) - \psi_{u+t}(A)}$$

Since Q_1 and Q_2 are both decreasing in t, we work with each separately. Precisely fix $\theta \in (0,1)$ and define Δ_1, Δ_2 as follows :

$$\Delta_1 \text{ the smallest positive number such that } Q_1(\Delta_1, B) \leqslant \theta$$

and

$$\Delta_2 \text{ the smallest positive number such that } Q_2(\Delta_2, B) \leqslant 1-\theta$$

Now take $\Delta = \Delta_1 + \Delta_2$, then $Q_1(\Delta, B) + Q_2(\Delta, B) \leqslant \theta + 1 - \theta = 1$. So this choice of Δ satisfies (4.3) and it remains now to estimate Δ_1 and Δ_2 separately.

4.7.3 Estimating Δ_1

We may write

$$Q_1(\Delta_1, B) = \left\| \sum_{i=1}^{n} \frac{B^{\frac{1}{2}} v_i v_i^t B^{\frac{1}{2}}}{u + \Delta_1 - \lambda_i} \right\|.$$

Put $\xi_i = B^{\frac{1}{2}} v_i v_i^t B^{\frac{1}{2}}$, $\mu_i = \psi(u - \lambda_i)$ and $\mu = \psi \Delta_1$. Denote P_S the orthogonal projection on $\text{span}(v_i)_{i \in S}$, clearly $rank(P_S) = |S|$. Then we have :

$$\begin{cases} \mathbb{E}\|\xi_i\| = 1 \\ \mathbb{P}\left(\left\| \sum_{i \in S} \xi_i \right\| \geqslant t \right) = \mathbb{P}\left(\|P_S B P_S\| \geqslant t \right) \leqslant \frac{c}{t^{1+\eta}} \quad \forall t \geqslant c\,|S| \\ \displaystyle\sum_{i=1}^{n} \frac{1}{\mu_i} = \frac{\psi_u(A)}{\psi} \leqslant 1 \\ \mu \text{ is the smallest positive number such that } \displaystyle\sum_{i=1}^{n} \frac{\xi_i}{\mu_i + \mu} \preceq \frac{\theta}{\psi} Id \end{cases}$$

Applying Lemma 4.20 below, we get $\mathbb{E}\mu \leqslant c(\eta) \left(\frac{\psi}{\theta}\right)^{1+\eta}$, so that

$$\mathbb{E}\Delta_1 \leqslant c(\eta) \frac{\psi^{\eta}}{\theta^{1+\eta}} \tag{4.4}$$

The Lemma we needed above is an analog of Lemma 3.5 appearing in [79]. We extend it to a matrix setting :

Lemma 4.20. *Suppose $\{\xi_i\}_{i \leqslant n}$ are symmetric positive semidefinite random matrices with* $\mathbb{E}\|\xi_i\| =$

1 *and*

$$\mathbb{P}\left(\left\|\sum_{i\in S}\xi_i\right\|\geqslant t\right)\leqslant \frac{c}{t^{1+\eta}}\quad \text{provided}\quad t>c|S|=c\sum_{i\in S}\mathbb{E}\|\xi_i\|,$$

for all subsets $S\subset[n]$ and some constants $c,\eta>0$. Consider positive numbers μ_i such that

$$\sum_{i=1}^{n}\frac{1}{\mu_i}\leqslant 1.$$

Let μ be the minimal positive number such that

$$\sum_{i=1}^{n}\frac{\xi_i}{\mu_i+\mu}\preceq K\cdot Id,$$

for some $K\geqslant C=4c$. Then $\mathbb{E}\mu\leqslant \frac{c(\eta)}{K^{1+\eta}}$.

Proof. For any $j\geqslant 0$, denote

$$I_j=\left\{i/\ 2^j\leqslant \mu_i<2^{j+1}\right\}$$

and let $n_j=|I_j|$. Note that

$$\sum_{j=0}^{+\infty}\frac{n_j}{2^{j+1}}\leqslant\sum_{j=0}^{+\infty}\sum_{i\in I_j}\frac{1}{\mu_i}=\sum_{i\leqslant n}\frac{1}{\mu_i}\leqslant 1$$

Define μ' as the minimal positive number such that

$$\forall j\geqslant 0,\qquad \frac{1}{2^j+\mu'}\left\|\sum_{i\in I_j}\xi_i\right\|\leqslant\varepsilon_j,$$

where $\varepsilon_j=\frac{K}{2}\frac{n_j}{2^{j+1}}\vee\frac{K}{2a}2^{-j\frac{\eta}{2+2\eta}}$ and $a=\sum_j 2^{-j\frac{\eta}{2+2\eta}}$. First note that $\mu\leqslant\mu'$; indeed,

$$\begin{aligned}\left\|\sum_{i=1}^{n}\frac{\xi_i}{\mu_i+\mu'}\right\|=\left\|\sum_{j=0}^{+\infty}\sum_{i\in I_j}\frac{\xi_i}{\mu_i+\mu'}\right\|&\leqslant\left\|\sum_{j=0}^{+\infty}\frac{1}{2^j+\mu'}\sum_{i\in I_j}\xi_i\right\|\\&\leqslant\sum_{j=0}^{+\infty}\frac{1}{2^j+\mu'}\left\|\sum_{i\in I_j}\xi_i\right\|\\&\leqslant\sum_{j=0}^{\infty}\varepsilon_j\\&\leqslant\frac{K}{2}+\frac{K}{2}=K,\end{aligned}$$

and since μ is the minimal positive number satisfying the inequality above, then $\mu\leqslant\mu'$. We may now estimate $\mathbb{E}\mu'$; to this aim, we need to look at $\mathbb{P}\{\mu' \geqslant t\}$. For $t \geqslant 0$,

$$\begin{aligned}\mathbb{P}\{\mu' \geqslant t\} &= \mathbb{P}\left\{\exists j \geqslant 0/ \ \frac{1}{2^j+t}\left\|\sum_{i\in I_j}\xi_i\right\| \geqslant \varepsilon_j\right\}\\ &\leqslant \sum_{j=0}^{+\infty}\mathbb{P}\left\{\frac{1}{2^j+t}\left\|\sum_{i\in I_j}\xi_i\right\| \geqslant \varepsilon_j\right\}\\ &= \sum_{j=0}^{+\infty}\mathbb{P}\left\{\left\|\sum_{i\in I_j}\xi_i\right\| \geqslant \varepsilon_j(2^j+t)\right\}\\ &\leqslant \sum_{j=0}^{+\infty}\frac{c}{[\varepsilon_j(2^j+t)]^{1+\eta}},\end{aligned}$$

where the last inequality comes from the fact that $\varepsilon_j(2^j+t) \geqslant \frac{K}{4}n_j \geqslant c|I_j|$ and by applying the hypothesis satisfied by the ξ_i. Now since $\varepsilon_j \geqslant \frac{K}{2a}2^{-j\frac{\eta}{2+2\eta}}$, we have

$$\mathbb{P}\{\mu' \geqslant t\} \leqslant \left(\frac{2a}{K}\right)^{1+\eta}\sum_{j=0}^{+\infty}\frac{c}{2^{-j\frac{\eta}{2}}(2^j+t)^{1+\eta}} \leqslant \frac{c(\eta)}{K^{1+\eta}}\sum_{j=0}^{+\infty}\frac{1}{(2^j+t)^{1+\frac{\eta}{2}}}$$

Now by integration we get,

$$\mathbb{E}\mu' = \int_0^{+\infty}\mathbb{P}\{\mu' \geqslant t\}\,dt \leqslant \frac{c(\eta)}{K^{1+\eta}}\sum_{j=0}^{+\infty}\int_0^{+\infty}\frac{1}{(2^j+t)^{1+\frac{\eta}{2}}}dt \leqslant \frac{c(\eta)}{K^{1+\eta}}$$

□

4.7.4 Estimating Δ_2

Suppose $\theta \leqslant \frac{1}{2}$. Since $\psi_u(A) - \psi_{u+t}(A) = t.\mathrm{Tr}\left((u.I_n - A)^{-1}\left((u+t).I_n - A\right)^{-1}\right)$ we can write

$$\begin{aligned}Q_2(t,B) = \frac{\sum_i \frac{\langle Bv_i,v_i\rangle}{(u+t-\lambda_i)^2}}{t\sum_i(u+t-\lambda_i)^{-1}(u-\lambda_i)^{-1}} &\leqslant \frac{1}{t}\frac{\sum_i \frac{\langle Bv_i,v_i\rangle}{(u+t-\lambda_i)(u-\lambda_i)}}{\sum_i(u+t-\lambda_i)^{-1}(u-\lambda_i)^{-1}}\\ &:= \frac{1}{t}P_2(t,B)\end{aligned} \tag{4.5}$$

First note that $P_2(t,B)$ can be written as $\sum_i \alpha_i(t)\langle Bv_i, v_i\rangle$ with $\sum_i \alpha_i(t) = 1$. Having this in mind, one can easily check that $\mathbb{E}P_2(t,B) = 1$ and

$$\mathbb{E}P_2(t,B)^{1+\frac{3\eta}{4}} \leqslant c(\eta), \tag{4.6}$$

where for the last inequality, we used the fact that B satisfies (MWR) with $p = 1 + \frac{3\eta}{4}$.
In order to estimate Δ_2, we will divide it into two parts as follows :

$$\Delta_2 = \Delta_2 \mathbf{1}_{\{P_2(0,B) \leqslant \frac{\theta}{4\psi}\}} + \Delta_2 \mathbf{1}_{\{P_2(0,B) > \frac{\theta}{4\psi}\}} := H_1 + H_2$$

Let us start by estimating $\mathbb{E}H_1$. Suppose that $P_2(0, B) \leqslant \frac{\theta}{4\psi}$ and denote

$$x = (1 + 4\theta)P_2(0, B).$$

Since $\psi_u(A) \leqslant \psi$, we have $(u - \lambda_i).\psi \geqslant 1$ $\forall i$ and therefore $u + x - \lambda_i \leqslant (1 + x\psi)(u - \lambda_i)$. This implies that

$$P_2(x, B) \leqslant (1 + x\psi)P_2(0, B).$$

Now write

$$Q_2(x, B) \leqslant \frac{1}{x} P_2(x, B) \leqslant \frac{1 + x\psi}{x} P_2(0, B) \leqslant \frac{1 + (1 + 4\theta)\frac{\theta}{4}}{1 + 4\theta} \leqslant 1 - \theta,$$

which means that

$$\Delta_2 \mathbf{1}_{\left\{P_2(0,B) \leqslant \frac{\theta}{4\psi}\right\}} \leqslant (1 + 4\theta)P_2(0, B)$$

and therefore

$$\mathbb{E}H_1 = \mathbb{E}\Delta_2 \mathbf{1}_{\left\{P_2(0,B) \leqslant \frac{\theta}{4\psi}\right\}} \leqslant 1 + 4\theta \tag{4.7}$$

Now it remains to estimate $\mathbb{E}H_2$. For that, we need to prove a moment estimate for Δ_2. First observe that using (4.6) we have

$$\mathbb{P}\{\Delta_2 > t\} = \mathbb{P}\{Q_2(t, B) > 1 - \theta\} \leqslant \mathbb{P}\{P_2(t, B) > t.(1 - \theta)\} \leqslant \frac{c(\eta)}{t^{1+\frac{3\eta}{4}}}$$

By integration, this implies

$$\mathbb{E}\Delta_2^{1+\frac{\eta}{2}} = \int_0^\infty \mathbb{P}\{\Delta_2 > t\}(1 + \frac{\eta}{2})t^{\frac{\eta}{2}} dt \leqslant \int_0^1 (1 + \frac{\eta}{2})t^{\frac{\eta}{2}} dt + \int_1^\infty \frac{c(\eta)}{t^{1+\frac{\eta}{4}}} \leqslant c(\eta)$$

Let $p' = 1 + \frac{\eta}{2}$, applying Hölder's inequality with $\frac{1}{p'} + \frac{1}{q'} = 1$ we have

$$\begin{aligned}\mathbb{E}H_2 = \mathbb{E}\Delta_2 \mathbf{1}_{\left\{P_2(0,B)>\frac{\theta}{4\psi}\right\}} &\leqslant \left(\mathbb{E}\Delta_2^{p'}\right)^{\frac{1}{p'}} \left(\mathbb{P}\left\{P_2(0,B) > \frac{\theta}{4\psi}\right\}\right)^{\frac{1}{q'}} \\ &\leqslant c(\eta)\left(\left(\frac{\psi}{\theta}\right)^{1+\frac{\eta}{2}} \mathbb{E}P_2(0,B)^{1+\frac{\eta}{2}}\right)^{\frac{1}{q'}} \\ &\leqslant c(\eta)\left(\frac{\psi}{\theta}\right)^{\frac{\eta}{2}}\end{aligned} \tag{4.8}$$

Looking at (4.7) and (4.8) we have

$$\mathbb{E}\Delta_2 \leqslant 1 + 4\theta + c(\eta)\left(\frac{\psi}{\theta}\right)^{\frac{\eta}{2}}$$

Putting the estimates of Δ_1 and Δ_2 together we deduce that

$$\mathbb{E}\Delta \leqslant 1 + 4\theta + c(\eta)\left(\frac{\psi}{\theta}\right)^{\frac{\eta}{2}} + c(\eta)\frac{\psi^\eta}{\theta^{1+\eta}}$$

We are now ready to finish the proof of Theorem 4.13. Take $u' = u + \Delta$, Δ being chosen as before with $\theta = \frac{\varepsilon}{8}$. Then taking $\psi = c(\eta)\varepsilon^{1+\frac{2}{\eta}}$ with the constant depending on η properly chosen, we get $\mathbb{E}\Delta \leqslant 1 + \varepsilon$.

Chapitre 5

log-concave matrices

A probability measure μ on $\mathbb{R}^n$ is called log-concave if for all $0 < t < 1$ and for all compact subsets $A, B \subset \mathbb{R}^n$ with positive measure one has

$$\mu\left((1-t)A + tB\right) \geqslant \mu(A)^{1-t}\mu(B)^t.$$

A random vector with a log-concave distribution is called log-concave. Borell ([16], [17]) characterized log-concave measures as follows : μ is a log-concave measure on $\mathbb{R}^n$ if and only if its density f with respect to the Lebesgue measure is log-concave i.e $\log(f)$ is a concave function.

Log-concave measures are a particular case of the class of convex measures introduced by Borell ([16], [17]). Let $s < 0$, a probability measure μ on $\mathbb{R}^n$ is called s-concave if for all $0 < t < 1$ and for all compact subsets $A, B \subset \mathbb{R}^n$ with positive measure one has

$$\mu\left((1-t)A + tB\right) \geqslant \left((1-t)\mu(A)^s + t\mu(B)^s\right)^{\frac{1}{s}}$$

A random vector with an s-concave distribution is called s-concave. When the support of an s-concave measure μ generates $\mathbb{R}^n$, a characterization of Borell ([16], [17]) states that μ is absolutely continuous with respect to the Lebesgue measure and its density h is of the form

$$h = f^{-\alpha} \quad \text{with} \quad \alpha = n - \frac{1}{s},$$

where $f : \mathbb{R}^n \longrightarrow (0, \infty]$ is a convex function. We will follow the notations used in [1], so we may take at each time a $(-\frac{1}{r})$-concave vector with $r > 0$ so that its density h is of the form

$h = f^{-(n+r)}$, where $f : \mathbb{R}^n \longrightarrow (0, \infty]$ is a convex function. Let us note that a log-concave vector is $(-\frac{1}{r})$-concave for any $r > 0$. Later we will define these notions for matrices ; the definition will be in a natural way since an $n \times m$ matrix can be considered as a vector in $\mathbb{R}^{nm}$. The difference point will be in the definition of the isotropic condition.

A random vector X in $\mathbb{R}^n$ is isotropic if $\mathbb{E}XX^t = Id$. An important fact is that a projection of an isotropic log-concave (resp. s-concave) vector is an isotropic log-concave (resp. s-concave) vector in the corresponding space.

Log-concave vectors were subject of many studies in recent years and the picture is more clear now. Let us state a fundamental result due to Paouris [59] :

Theorem 5.1 (Paouris)**.** *Let X be an isotropic log-concave vector in $\mathbb{R}^n$. For every $t \geqslant 1$, we have*

$$\mathbb{P}\left\{\|X\|_2 \geqslant ct\sqrt{n}\right\} \leqslant \exp(-t\sqrt{n})$$

Another important result is the small ball probability estimate for isotropic log-concave vectors obtained by Paouris [60]. We do not state the result in its full generality, we refer to [60] for the exact statement.

Theorem 5.2 (Paouris)**.** *Let X be an isotropic log-concave vector in $\mathbb{R}^n$. For every $\varepsilon \in (0, 1)$,*

$$\mathbb{P}\left\{\|X\|_2 \leqslant c\varepsilon\sqrt{n}\right\} \leqslant \varepsilon^{c\sqrt{n}}$$

In [2], a new short proof was given to Paouris's result and it is shown in [1] that the same techniques extend to the case of convex measures ; a large deviation inequality and a small ball probability estimate were given in [1] for $(-\frac{1}{r})$-concave isotropic random vectors. Let us start with the large deviation inequality :

Theorem 5.3 (AGLLOPT)**.** *Let $r > 2$ and let $X \in \mathbb{R}^n$ be a $(-\frac{1}{r})$-concave isotropic random vector. Then for every $t > 0$,*

$$\mathbb{P}\left\{\|X\|_2 \geqslant t\sqrt{n}\right\} \leqslant \left(\frac{c\max\{1, r/\sqrt{n}\}}{t}\right)^{\frac{r}{2}}.$$

In particular, if $r \geqslant 2\sqrt{n}$, then for every $6c \leqslant t \leqslant 3cr/\sqrt{n}$,

$$\mathbb{P}\left\{\|X\|_2 \geqslant t\sqrt{n}\right\} \leqslant \exp(-c_0t\sqrt{n}),$$

where c and c_0 are universal positive constants.

We may now turn to the small ball probability estimate obtained in [1]. We do not state the result in its full generality, we refer to [1] for the exact statement

Theorem 5.4 (AGLLOPT)**.** *Let $r > 2$ and let $X \in \mathbb{R}^n$ be a $(-\frac{1}{r})$-concave isotropic random vector. Denote $\alpha = \min\{r, \sqrt{n}\}$, then for every $\varepsilon \in (0,1)$,*

$$\mathbb{P}\left\{\|X\|_2 \leqslant c\varepsilon\sqrt{n}\right\} \leqslant \varepsilon^{c\alpha}$$

where c is a universal positive constant.

Back to the log-concave case, Paouris's results mean that the euclidean norm of an isotropic log-concave vector in $\mathbb{R}^n$ lies, up to a constant, at the level $\sqrt{n}$ with high probability. This can be refined to show that with high probability, the equivalence constant between the euclidean norm of an isotropic log-concave vector in $\mathbb{R}^n$ and $\sqrt{n}$, is almost 1. This is called the thin-shell estimate ; let us state the result obtained by Guédon-Milman [36] :

Theorem 5.5 (Guédon-Milman)**.** *Let X denote an isotropic log-concave vector in $\mathbb{R}^n$. Then,*

$$\mathbb{P}\left\{\left|\|X\|_2 - \sqrt{n}\right| \geqslant t\sqrt{n}\right\} \leqslant C\exp\left(-c\sqrt{n}\min(t, t^3)\right), \quad \forall t \geqslant 0.$$

5.1 Isotropic log-concave matrices

A natural way to define a log-concave matrix is to ask that it has a log-concave distribution. However, for the isotropic condition we will define it in a coherent way with what is done in the previous chapter.

Definition 5.6. *Let A be an $n \times m$ random matrix and denote by $(C_i)_{i \leqslant m}$ its columns. We will say that A is an isotropic log-concave matrix if $A'^t = \sqrt{m}(C_1^t, .., C_m^t)$ is an isotropic log-concave random vector in $\mathbb{R}^{nm}$.*

Remark 5.7. *Let $(a_{i,j})$ the entries of A. Saying that A' is isotropic means that*

$$\mathbb{E}a_{i,j}a_{k,l} = \frac{1}{m}\delta_{(i,j),(k,l)}$$

This implies that for any $n \times m$ matrix M we have

$$\mathbb{E}\left\langle A, M\right\rangle A = \mathbb{E}\mathrm{Tr}\left(A^tM\right)A = \frac{1}{m}M.$$

One can view this as an analogue to the isotropic condition in the vector case : in fact if $A = X$ is a vector (i.e an $n \times 1$ matrix), the above condition would be

$$\mathbb{E} \langle X, y \rangle X = y \quad \text{for all } y \in \mathbb{R}^n,$$

which means that X is isotropic in $\mathbb{R}^n$.

Let us now write the large deviation inequality and the small ball probability estimate satisfied by an isotropic log-concave matrix.

Proposition 5.8. *Let A be an $n \times m$ isotropic log-concave matrix and denote $B = AA^t$. Then for every orthogonal projection P on $\mathbb{R}^n$ we have the following large deviation estimate for* $\mathrm{Tr}(PB)$

$$\mathbb{P}\left\{\mathrm{Tr}(PB) \geqslant c_1 t\right\} \leqslant \exp\left(-\sqrt{t.m}\right) \quad \forall t \geqslant rank(P) \tag{5.1}$$

and a small ball probability estimate

$$\mathbb{P}\left\{\mathrm{Tr}(PB) \geqslant c_2 \varepsilon.rank(P)\right\} \leqslant \varepsilon^{c_2\sqrt{m.rank(P)}} \quad \forall \varepsilon \leqslant 1. \tag{5.2}$$

Moreover, we also have a thin-shell estimate

$$\mathbb{P}\left\{|\mathrm{Tr}(PB) - \mathrm{rank}(P)| \geqslant t.\mathrm{rank}(P)\right\} \leqslant C\exp\left(-ct^3\sqrt{m.\mathrm{rank}(P)}\right) \quad \forall t \leqslant 1. \tag{5.3}$$

Proof. Let P be an orthogonal projection on $\mathbb{R}^n$ and denote $P' = I_m \otimes P$. As we have seen before $\mathrm{Tr}(PB) = \|PA\|_{\mathrm{HS}}^2 = \frac{1}{m}\|P'A'\|_2^2$ and $rank(P') = m.rank(P)$. Since $P'A'$ is an isotropic log-concave vector, then using Theorem 5.1 for $P'A'$, we have

$$\mathbb{P}\left\{\|P'A'\|_2^2 \geqslant c_1 u\right\} \leqslant \exp\left(-\sqrt{u}\right) \quad \forall u \geqslant rank(P').$$

Let $t \geqslant rank(P)$ and write $u = t.m$. Since $u \geqslant m.rank(P) = rank(P')$ we have

$$\mathbb{P}\left\{m.Tr(PB) \geqslant c_1 t.m\right\} \leqslant \exp\left(-\sqrt{t.m}\right)$$

which gives the large deviation estimate stated above.

For the small ball probability estimate, we apply Theorem 5.2 to $P'A'$:

$$\mathbb{P}\left\{\|P'A'\|_2^2 \geqslant c_2 \varepsilon.rank(P')\right\} \leqslant \varepsilon^{c_2\sqrt{rank(P')}} \quad \forall \varepsilon \leqslant 1.$$

Writing this in terms of B and P, we easily get the conclusion. Using Theorem 5.5 with the same procedure as above, we get the thin-shell estimate. $\square$

In [79], it was shown that an isotropic log-concave vector satisfies (SR) and in the previous chapter we showed in Proposition 4.6 how to pass from (SR) to $(MSR*)$. Therefore, we may apply Theorem 4.3 to log-concave matrices and get the following :

Proposition 5.9. *Let A be an $n \times m$ isotropic log-concave matrix. Then $B = AA^t$ satisfies (MSR). Moreover $\forall \varepsilon > 0$, taking $N > c(\varepsilon)n$ independent copies of B we have*

$$\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N} B_i - I_n\right\| \leqslant \varepsilon,$$

where $c(\varepsilon) = c\varepsilon^{-2-o(1)}$.

Proof. Note first that since A is isotropic in the sense of definition 5.6, then $B = AA^t$ satisfies $\mathbb{E}B = I_n$.

By proposition 5.8, B satisfies

$$\mathbb{P}\left(\mathrm{Tr}(PB) \geqslant c_1 t\right) \leqslant \exp(-\sqrt{tm}) \quad \forall t \geqslant rank(P) \text{ and } \forall P \text{ orthogonal projection of } \mathbb{R}^n.$$

and therefore $(MSR*)$. Applying theorem 4.3 we deduce the result. $\square$

5.2 Eigenvalues of the empirical sum of a log-concave matrix

The probability estimate for these log-concave matrices are strong enough to allow us obtain some results with high probability rather than in expectation as was the case before. Precisely, we can prove the following :

Theorem 5.10. *Let n, m and N some fixed integers. Let A be an $n \times m$ isotropic log-concave matrix and denote $B = AA^t$. For any $\varepsilon \in (0,1)$, if $m \geqslant \frac{C}{\varepsilon^6}\left[\log(CnN)\right]^2$, then with probability $\geqslant 1 - \exp(-c\varepsilon^3\sqrt{m})$ we have*

$$\lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \leqslant 1 + \varepsilon + \frac{6n}{\varepsilon N},$$

where $(B_i)_{i\leqslant N}$ are independent copies of B.

Proof. The proof of Theorem 5.10 follows the same ideas as in the previous chapter. Let $\varepsilon \in (0,1)$, we only need the following property satisfied by our matrix $B = AA^t$:

$$\mathbb{P}\left(\langle Bx, x\rangle \geqslant 1+\frac{\varepsilon}{2}\right) \leqslant \exp(-c\varepsilon^3\sqrt{m}) \quad \forall x \in \mathbb{S}^{n-1}.$$

This is obtained by applying (5.3) for rank 1 projections and looking only at the large deviation part.

Define Δ, ψ and α as follows :

$$\Delta = 1+\varepsilon; \quad \psi = \frac{\varepsilon}{6} \quad \text{and} \quad \alpha = \frac{1+\frac{\varepsilon}{2}}{1+\varepsilon}.$$

Recall some notations

$$A_0 = 0,\ A_1 = B_1,\ A_2 = A_1 + B_1,\ ..,\ A_N = A_{N-1} + B_N = \sum_{i=1}^{N} B_i.$$

Denote

$$u_0 = \frac{n}{\psi},\ u_1 = u_0 + \Delta,\ u_2 = u_1 + \Delta,\ ..,\ u_N = u_{N-1} + \Delta = (1+\varepsilon)N + \frac{6n}{\varepsilon}.$$

Define

$$\psi_{u_i}(A_i) = \mathrm{Tr}\left(u_i.I_n - A_i\right)^{-1},$$

the corresponding potential function when $A_i \prec u_i.I_n$. Denote by $\Im_i$ the event

$$\Im_i := \text{“}A_i \prec u_i.I_n \quad \text{and} \quad \psi_{u_i}(A_i) \leqslant \psi\text{”}.$$

Clearly $\mathbb{P}\left(\Im_0\right) = 1$. Suppose now that $\Im_i$ is satisfied ; as we have seen in Lemma 4.19, the following condition is sufficient for the occurrence of the event $\Im_{i+1}$:

$$Q_2(\Delta, B_{i+1}) + Q_1(\Delta, B_{i+1}) \leqslant 1$$

Note that

$$Q_2(\Delta, B_{i+1}) \leqslant \frac{1}{\Delta} P_2(\Delta, B_{i+1}),$$

where P_2 is defined in (4.5). Now denoting λ_j the eigenvalues of A_i and v_j the corresponding eigenvectors, taking the probability with respect to B_{i+1} one can write

$$
\begin{aligned}
&\mathbb{P}\left(Q_2(\Delta, B_{i+1}) + Q_1(\Delta, B_{i+1}) > 1\right) \leqslant \mathbb{P}\left(\frac{1}{\Delta}P_2(\Delta, B_{i+1}) + Q_1(\Delta, B_{i+1}) > 1\right)\\
&\leqslant \mathbb{P}\left(\frac{1}{\Delta}P_2(\Delta, B_{i+1}) > \alpha\right) + \mathbb{P}\left(Q_1(\Delta, B_{i+1}) > 1-\alpha\right)\\
&\leqslant \mathbb{P}\left(\sum_{j=1}^{n} \frac{\langle B_{i+1}v_j, v_j\rangle}{(u_{i+1}-\lambda_j)(u_i-\lambda_j)} > \alpha\Delta \sum_{j=1}^{n} \frac{1}{(u_{i+1}-\lambda_j)(u_i-\lambda_j)}\right)\\
&\quad + \mathbb{P}\left(\sum_{j=1}^{n} \frac{\langle B_{i+1}v_j, v_j\rangle}{u_{i+1}-\lambda_j} > 1-\alpha\right)\\
&\leqslant \mathbb{P}\left(\exists j \leqslant n/\ \ \langle B_{i+1}v_j, v_j\rangle > 1+\frac{\varepsilon}{2}\right) + \mathbb{P}\left(\exists j \leqslant n/\ \ \langle B_{i+1}v_j, v_j\rangle > \frac{1-\alpha}{\psi}\right)\\
&\leqslant 2Cn.\exp(-c\varepsilon^3\sqrt{m})
\end{aligned}
$$

So we have shown that $\mathbb{P}\left(\Im_{i+1}|\Im_i\right) \geqslant 1 - 2Cn.\exp(-c\varepsilon^3\sqrt{m})$. Since B_i are independent we have

$$
\begin{aligned}
\mathbb{P}\left(\lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \leqslant 1+\varepsilon+\frac{6n}{\varepsilon N}\right) &\geqslant \mathbb{P}\left(\Im_N\right)\\
&\geqslant \mathbb{P}\left(\Im_N|\Im_{N-1}\right)\mathbb{P}\left(\Im_{N-1}|\Im_{N-2}\right)..\mathbb{P}\left(\Im_0\right)\\
&\geqslant 1 - 2CNn.\exp(-c\varepsilon^3\sqrt{m})
\end{aligned}
$$

Therefore, Theorem 5.10 follows by the choice of m. □

Remark 5.11. *Note that in the previous proof, we only used the large deviation inequality given by the thin-shell estimate (5.3). If one uses the deviation inequality given by (5.1), then by the same proof it can be proved that*

$$
\lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \leqslant C(1+\frac{n}{N}),
$$

with high probability and with similar condition on m. The advantage of using thin-shell is that we can get an estimate close to 1.

By the same techniques, we also get an estimate of the smallest eigenvalue.

Theorem 5.12. *Let n, m and N some fixed integers. Let A be an $n\times m$ isotropic log-concave matrix and denote $B = AA^t$. For any $\varepsilon \in (0,1)$, if $m \geqslant \frac{C}{\varepsilon^6}\left[\log(CnN)\right]^2$, then with probability*

$\geqslant 1 - \exp(-c\varepsilon^3\sqrt{m})$ *we have*

$$\lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \geqslant 1 - \varepsilon - \frac{3n}{\varepsilon N},$$

where $(B_i)_{i\leqslant N}$ *are independent copies of* B.

Proof. Here we will use the lower and the upper estimate given by thin-shell (5.3). Applying (5.3) for rank 1 projections, we have :

$$\mathbb{P}\left(1 - \frac{\varepsilon}{2} \leqslant \langle Bx, x\rangle \leqslant 1 + \frac{\varepsilon}{2}\right) \geqslant 1 - C\exp(-c\varepsilon^3\sqrt{m}) \quad \forall x \in \mathbb{S}^{n-1}.$$

Define δ, ϕ and α as follows :

$$\delta = 1 - \varepsilon; \quad \phi = \frac{\varepsilon}{3} \quad \text{and} \quad \alpha = \frac{1 - \frac{\varepsilon}{2}}{1 - \varepsilon}.$$

Recall some notations

$$A_0 = 0,\ A_1 = B_1,\ A_2 = A_1 + B_1,\ ..,\ A_N = A_{N-1} + B_N = \sum_{i=1}^{N} B_i.$$

Denote

$$l_0 = -\frac{n}{\phi},\ l_1 = l_0 + \delta,\ l_2 = l_1 + \delta,\ ..,\ l_N = l_{N-1} + \delta = N(1-\varepsilon) - \frac{3n}{\varepsilon}.$$

Define

$$\phi_{l_i}(A_i) = \mathrm{Tr}\left(A_i - l_i.I_n\right)^{-1},$$

the corresponding potential function when $A_i \succeq l_i.I_n$. Note also that $\delta \leqslant \frac{1}{\phi}$.
Denote by $\Im_i$ the event

$$\Im_i := \text{“}A_i \succeq l_i.I_n \quad \text{and} \quad \phi_{l_i}(A_i) \leqslant \phi\text{”}.$$

Clearly $\mathbb{P}\left(\Im_0\right) = 1$. Suppose now that $\Im_i$ is satisfied, following what was done after Lemma 4.15, condition (4.2) is sufficient for the occurrence of the event $\Im_{i+1}$:

$$\frac{1}{\delta}q_2(\delta, B_{i+1}) - q_1(\delta, B_{i+1}) \geqslant 1$$

Denoting λ_j the eigenvalues of A_i and v_j the corresponding eigenvectors, taking the probability with respect to B_{i+1} one can write :

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{\delta}q_2(\delta, B_{i+1}) - q_1(\delta, B_{i+1}) < 1\right) \leqslant \\
&\leqslant \mathbb{P}\left(\frac{1}{\delta}q_2(\delta, B_{i+1}) < \alpha\right) + \mathbb{P}\left(q_1(\delta, B_{i+1}) > \alpha - 1\right) \\
&\leqslant \mathbb{P}\left(\sum_{j=1}^{n} \frac{\langle B_{i+1}v_j, v_j\rangle}{(\lambda_j - l_{i+1})^2} < \alpha\delta \sum_{j=1}^{n} \frac{1}{(\lambda_j - l_{i+1})^2}\right) + \mathbb{P}\left(\sum_{j=1}^{n} \frac{\langle B_{i+1}v_j, v_j\rangle}{\lambda_j - l_{i+1}} > \alpha - 1\right) \\
&\leqslant \mathbb{P}\left(\exists j \leqslant n/ \ \langle B_{i+1}v_j, v_j\rangle < 1 - \frac{\varepsilon}{2}\right) + \mathbb{P}\left(\exists j \leqslant n/ \ \langle B_{i+1}v_j, v_j\rangle > \frac{\alpha - 1}{\phi}\right) \\
&\leqslant 2Cn.\exp(-c\varepsilon^3\sqrt{m})
\end{aligned}$$

So we have shown that $\mathbb{P}\left(\Im_{i+1}|\Im_i\right) \geqslant 1 - 2Cn.\exp(-c\varepsilon^3\sqrt{m})$. Since B_i are independent we have :

$$\begin{aligned}
\mathbb{P}\left(\lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \geqslant 1 - \varepsilon - \frac{3n}{\varepsilon N}\right) &\geqslant \mathbb{P}\left(\Im_N\right) \\
&\geqslant \mathbb{P}\left(\Im_N|\Im_{N-1}\right)\mathbb{P}\left(\Im_{N-1}|\Im_{N-2}\right)..\mathbb{P}\left(\Im_0\right) \\
&\geqslant 1 - 2CNn.\exp(-c\varepsilon^3\sqrt{m})
\end{aligned}$$

Therefore, Theorem 5.12 follows by the choice of m.

□

Remark 5.13. *Note that in the previous proof, we used the large deviation inequality alongside the small ball probability estimate given by thin-shell (5.3). If one uses the deviation inequality given by (5.1) alongside the small ball probability estimate given by (5.2), then by the same proof it can be proved that*

$$\lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \leqslant -c + C\frac{n}{N},$$

with high probability and with similar condition on m. The advantage of using thin-shell is that we can get an estimate close to 1.

Combining the two previous results, we will be able to obtain, with high probability, a similar result to Proposition 5.9 for log-concave matrices :

Theorem 5.14. *Let A be an $n \times m$ isotropic log-concave matrix. For any $\varepsilon \in (0,1)$, if $m \geqslant \frac{C}{\varepsilon^6}\left[\log(\frac{Cn}{\varepsilon})\right]^2$, taking*

$$N \geqslant \frac{96n}{\varepsilon^2}$$

copies $A_1, ..., A_N$ of A, then with probability $\geqslant 1 - \exp(-c\varepsilon^3\sqrt{m})$ we have

$$\left\| \frac{1}{N}\sum_{i=1}^{N} A_i A_i^t - Id \right\| \leqslant \varepsilon$$

Proof. Let $\varepsilon \in (0,1)$, $N \geqslant \frac{6n}{\varepsilon^2}$ and suppose that m satisfies the assumption of the theorem. Note that

$$\mathbb{P}\left\{\left\| \frac{1}{N}\sum_{i=1}^{N} A_i A_i^t - Id \right\| > 4\varepsilon\right\} \leqslant \mathbb{P}\left\{\lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} A_i A_i^t\right) > 1 + 2\varepsilon\right\} + \mathbb{P}\left\{\lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^{N} A_i A_i^t\right) < 1 - 2\varepsilon\right\}$$

and therefore it is sufficient to apply Theorem 5.10 and Theorem 5.12. □

5.3 Concrete examples of isotropic log-concave matrices

To find a log-concave matrix, we need to define a log-concave function over $\mathbb{M}_{n,m}$. It is indeed sufficient to search for convex functions and then take the exponential of the opposite function. A natural way to define a function over $\mathbb{M}_{n,m}$ is to define one over the singular values. To simplify the idea, let us for instance give an example on $\mathbb{SM}^n$ the space of symmetric matrices ; if A is an $n \times n$ symmetric matrix and $V : \mathbb{R} \longrightarrow \mathbb{R}$ a function, then denoting λ_i the eigenvalues of A one can define

$$\hat{V}(A) = \sum_{j \leqslant n} V(\lambda_i)$$

This definition has the advantage of transferring properties of V to $\hat{V}$. Let us illustrate this fact in the following proposition :

Proposition 5.15. *Let $V : \mathbb{R} \longrightarrow \mathbb{R}$ be a convex function. Define $\hat{V}$ a function on $\mathbb{SM}^n$ by*

$$\hat{V}(A) = \sum_{i=1}^{n} V\left(\lambda_i(A)\right)$$

Then for any $A \in \mathbb{SM}^n$,

$$\hat{V}(A) = \sup \sum_{i=1}^{n} V\left(\langle Ae_i, e_i\rangle\right),$$

the supremum being taken over all orthonormal basis of $\mathbb{R}^n$*. Moreover* $\hat{V}$ *is a convex function on* $\mathbb{SM}^n$*.*

Proof. Since A is symmetric, apply the spectral theorem to get v_i an orthonormal basis of eigenvectors associated to $\lambda_i(A)$. Since $\lambda_i(A) = \langle Av_i, v_i\rangle$, then clearly $\hat{V}(A) \leqslant \sup \sum_{i=1}^{n} V\left(\langle Ae_i, e_i\rangle\right)$. Let $(e_i)_{i\leqslant n}$ be an orthonormal basis on $\mathbb{R}^n$. Write $Ae_i = \sum_{j=1}^{n} \langle Ae_i, v_j\rangle\, v_j = \sum_{j=1}^{n} \lambda_j \langle e_i, v_j\rangle\, v_j$. Then

$$\begin{aligned}
\sum_{i=1}^{n} V\left(\langle Ae_i, e_i\rangle\right) &= \sum_{i=1}^{n} V\left(\sum_{j=1}^{n} \lambda_j \langle e_i, v_j\rangle^2\right) \\
&\leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} \langle e_i, v_j\rangle^2 V(\lambda_j) \quad \text{since } V \text{ convex} \\
&\leqslant \sum_{j=1}^{n} V(\lambda_j),
\end{aligned}$$

which prove the second inequality. Now to prove that $\hat{V}$ is convex, take $A, B \in \mathbb{SM}^n$ and $\theta \in (0,1)$ and write

$$\begin{aligned}
\hat{V}\left(\theta A + (1-\theta)B\right) &= \sup \sum_{i=1}^{n} V\left(\theta\langle Ae_i, e_i\rangle + (1-\theta)\langle Be_i, e_i\rangle\right) \\
&\leqslant \sup\left[\theta \sum_{i=1}^{n} V\left(\langle Ae_i, e_i\rangle\right) + (1-\theta)\sum_{i=1}^{n} V\left(\langle Ae_i, e_i\rangle\right)\right] \\
&\leqslant \theta\hat{V}(A) + (1-\theta)\hat{V}(B)
\end{aligned}$$

□

It turns out that this phenomena is more general as is shown in [47]. For $x \in \mathbb{R}^k$, we denote by $\hat{x}$ the vector with components $|x_i|$ arranged in nonincreasing order. Let $f : \mathbb{R}^k \longrightarrow \mathbb{R}$, we say that f is absolutely symmetric if $f(x) = f(\hat{x})$ for all $x \in \mathbb{R}^k$. (For example, $\|\cdot\|_p$ is absolutely symmetric).

Define F a function on $\mathbb{M}_{n,m}$ by $F(A) = f\left(s_1(A), .., s_k(A)\right)$ for $A \in \mathbb{M}_{n,m}$ and $k = \min(n,m)$. It was shown by Lewis [47] that f is absolutely symmetric if and only if F is unitary invariant and of this form. Moreover, f is convex if and only if F is convex.

Let A be an $n \times m$ random matrix whose density with respect to Lebesgue measure is given by $G(A) = \exp\left(-f\left(s_1(A), .., s_k(A)\right)\right)$, where f is an absolutely symmetric convex function. By the remark above, G is log-concave. This covers the case of random matrices with density of the form $\exp\left(-\sum_i V\left(s_i(A)\right)\right)$, where V is an increasing convex function on $\mathbb{R}^+$. When $V(x) = x^2$, this would be the gaussian unitary ensemble GUE.
Let $(a_{i,j})$ be the entries of A. By a good normalization of f, we can suppose that A satisfies

$$\mathbb{E} a_{i,j} a_{k,l} = \frac{1}{m} \delta_{(i,j),(k,l)}$$

To see this, fix (i,j) and (k,l) two different indices. Note $D_j = diag(1, .., -1, .., 1)$ the $m \times m$ diagonal matrix where the -1 is on the j^{th} term. Let $E_{(i,k)}$ be the $n \times n$ matrix obtained by swapping the i^{th} and k^{th} rows in the identity matrix. Note also $F_{(j,l)}$ the $m \times m$ matrix obtained by swapping the j^{th} and l^{th} rows in the identity matrix.

It is easy to see that AD_j change the j^{th} column of A to its opposite and keep the rest unchanged. Note that AD_j has the same singular values as A.

Similarly, $E_{(i,k)} A F_{(j,l)}$ permute $a_{i,j}$ with $a_{k,l}$ and keep the other terms unchanged. Note also that $E_{(i,k)} A F_{(j,l)}$ has the same singular values as A.

Finally note that these two transformations has a Jacobian equal to 1, and since f is absolutely symmetric, these transformations, which preserve the singular values, don't affect the density.
If $j \neq l$, by a change of variables $M = AD_j$ the density is invariant and we have

$$\int a_{i,j} a_{k,l} G(A) dA = -\int a_{i,j} a_{k,l} G(A) dA$$

Doing the change of variables $M = D_i A$ when $i \neq k$, we can conclude that

$$\mathbb{E} a_{i,j} a_{k,l} = 0 \quad \text{ if } (i,j) \neq (k,l)$$

Now by a change of variables $M = E_{(i,j)} A F_{(k,l)}$ the density is invariant and we have

$$\int a_{i,j}^2 G(A) dA = \int a_{k,l}^2 G(A) dA$$

This implies that

$$\int a_{i,j}^2 G(A)dA = \frac{1}{nm} \sum_{k\leqslant n, l\leqslant m} \int a_{k,l}^2 G(A)dA = \frac{1}{nm} \int \|A\|_{\mathrm{HS}}^2 G(A)dA$$

Now we may normalize f in order to make the previous term equal $\frac{1}{m}$. Suppose that

$$\frac{1}{n} \int \|A\|_{\mathrm{HS}}^2 G(A)dA = c$$

Define $\hat{f}(x) = f(\sqrt{c}x) - nm\log(\sqrt{c})$ and $\hat{G}(A) = \exp\left(-\hat{f}(s_1(A),..,s_k(A))\right)$.

Note that $\hat{G}$ is a probability density. Indeed, by the change of variables $M = \sqrt{c}A$ we have

$$\begin{aligned}\int \hat{G}(A)dA &= \int \exp\left(-f(\sqrt{c}s_1(A),..,\sqrt{c}s_k(A))\right)(\sqrt{c})^{nm}dA \\ &= \int \exp\left(-f(s_1(M),..,s_k(M))\right)dM = 1\end{aligned}$$

Note also that $\hat{G}$ satisfies our isotropic condition. Indeed, by the same change of variables we can write

$$\frac{1}{n} \int \|A\|_{\mathrm{HS}}^2 \hat{G}(A)dA = \frac{1}{cn} \int \|M\|_{\mathrm{HS}}^2 G(M)dM = 1$$

As a conclusion, we can deduce that such matrices are isotropic log-concave. Moreover, since in this case $\mathbb{E}A^tA = \frac{n}{m}I_m$, then we also have that $\sqrt{\frac{m}{n}}A^t$ is an $m \times n$ isotropic log-concave matrix. We summarize this in the following proposition :

Proposition 5.16. *Let A be an $n \times m$ random matrix whose density with respect to Lebesgue is given by*

$$G(A) = \exp\left(-f(s_1(A),...,s_k(A))\right),$$

where f is an absolutely symmetric convex function, properly normalized as above and $k = \min(n,m)$. Then A is an isotropic log-concave matrix, and $\sqrt{\frac{n}{m}}A^t$ is an $m \times n$ isotropic log-concave matrix.

Applying Theorem 5.10 and Theorem 5.12 for A and A^t we get :

Proposition 5.17. *Let A be an $n \times m$ random matrix whose density with respect to Lebesgue is given by*

$$G(A) = \exp\left(-f(s_1(A),...,s_k(A))\right),$$

where f is an absolutely symmetric convex function, properly normalized as above and $k = \min(n,m)$.
Suppose that $m \geqslant \frac{C}{\varepsilon^6}\left[\log(\frac{Cn}{\varepsilon})\right]^2$ and $n \geqslant \frac{C}{\varepsilon^6}\left[\log(\frac{Cm}{\varepsilon})\right]^2$, taking $N = \frac{96\max(n,m)}{\varepsilon^2}$ then with probability $\geqslant 1-\exp(-c\varepsilon^3\sqrt{k})$ we have

$$1-\varepsilon \leqslant \lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^{N} A_iA_i^t\right) \leqslant \lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} A_iA_i^t\right) \leqslant 1+\varepsilon$$

and

$$(1-\varepsilon)\frac{n}{m} \leqslant \lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^{N} A_i^tA_i\right) \leqslant \lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N} A_i^tA_i\right) \leqslant (1+\varepsilon)\frac{n}{m}$$

5.4 Isotropic $(-\frac{1}{r})$-concave matrices

As we have done previously for log-concave matrices, we will start by defining isotropic $(-\frac{1}{r})$-concave matrices.

Definition 5.18. *Let A be an $n \times m$ random matrix and denote by $(C_i)_{i\leqslant m}$ its columns. We will say that A is an isotropic $(-\frac{1}{r})$-concave matrix if $A'^t = \sqrt{m}(C_1^t, .., C_m^t)$ is an isotropic $(-\frac{1}{r})$-concave random vector in $\mathbb{R}^{nm}$.*

Let us now write the large deviation inequality satisfied by an isotropic $(-\frac{1}{r})$-concave matrix. This can be obtained as in Proposition 5.8, by using the large deviation inequality given in Theorem 5.3.

Proposition 5.19. *Let $r > 2$ and A an $n \times m$ isotropic $(-\frac{1}{r})$-concave matrix and denote $B = AA^t$. For every orthogonal projection P on $\mathbb{R}^n$ we have the following large deviation estimate for $\mathrm{Tr}(PB)$; for every $t > 0$,*

$$\mathbb{P}\left\{\mathrm{Tr}(PB) \geqslant t\sqrt{\mathrm{rank}(P)}\right\} \leqslant \left(\frac{c\max\{1, r/\sqrt{m.\mathrm{rank}(P)}\}}{t}\right)^{\frac{r}{2}}.$$

In particular, if $r \geqslant 2\sqrt{m.\mathrm{rank}(P)}$, then for every $6c \leqslant t \leqslant 3cr/\sqrt{m.\mathrm{rank}(P)}$,

$$\mathbb{P}\left\{\mathrm{Tr}(PB) \geqslant t\sqrt{\mathrm{rank}(P)}\right\} \leqslant \exp\left(-c_0 t\sqrt{m.\mathrm{rank}(P)}\right),$$

where c and c_0 are universal positive constants.

In a similar way, we have also a small ball probability estimate.

Proposition 5.20. *Let $r > 2$ and A an $n \times m$ isotropic $(-\frac{1}{r})$-concave matrix and denote $B = AA^t$. For every orthogonal projection P on $\mathbb{R}^n$ and every $\varepsilon \in (0,1)$,*

$$\mathbb{P}\left\{\operatorname{Tr}(PB) \leqslant c\varepsilon\sqrt{\operatorname{rank}(P)}\right\} \leqslant \varepsilon^{c\alpha},$$

where $\alpha = \min\{r, \sqrt{m.\operatorname{rank}(P)}\}$ and c is a universal positive constant.

In [1], it was shown that, with r properly chosen, an isotropic $(-\frac{1}{r})$-concave random vector satisfies (SR). In the previous chapter we showed in Proposition 4.6 how to pass from (SR) to $(MSR*)$. Therefore, we may apply Theorem 4.3 to isotropic $(-\frac{1}{r})$-concave matrices and get the following :

Proposition 5.21. *Let $a > 0$ and $r = \max\{4, 2a\log n\}$. Let $A_1, ..., A_N$ be independent isotropic $(-1/r)$-concave $n \times m$ matrices. Then for every $\varepsilon \in (0,1)$ and every $N \geqslant C(\varepsilon, a)n$, one has*

$$\mathbb{E}\left\|\frac{1}{N}\sum_{j\leqslant N} A_jA_j^t - Id\right\| \leqslant \varepsilon,$$

where $C(\varepsilon, a)$ depends only on a and ε.

5.5 Eigenvalues of the empirical sum of a $(-\frac{1}{r})$-concave matrix

As for log-concave matrices, using the same techniques as before, we will derive some estimates on the smallest and largest eigenvalues of the empirical sum of a $(-\frac{1}{r})$-concave matrix. Let us start with the estimate of the largest eigenvalue :

Proposition 5.22. *Let n, m and N some fixed integers. Let $r > 2$ and A be an $n \times m$ isotropic $(-\frac{1}{r})$-concave matrix and denote $B = AA^t$. If*

$$\min\left\{r, \sqrt{m}\right\} \geqslant C\log(2nN), \tag{5.4}$$

then with probability $\geqslant 1 - \exp\left(-c\min\{r, \sqrt{m}\}\right)$ we have

$$\lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^N B_i\right) \leqslant C(1 + \frac{n}{N}),$$

where $(B_i)_{i\leqslant N}$ are independent copies of B and c, C are universal positive constants.

Proof. Let us denote $\alpha = \min\{r, \sqrt{m}\}$. The proof of Proposition 5.22 follows the same procedure as in the proof of Theorem 5.10. We only need the following property satisfied by our matrix $B = AA^t$:

$$\mathbb{P}\left(\langle Bx, x\rangle \geqslant C\right) \leqslant \exp(-c\alpha) \quad \forall x \in \mathbb{S}^{n-1}.$$

This is obtained by applying Proposition 5.19 for rank 1 projections. Define Δ and ψ as follows :

$$\Delta = 2C \quad \text{ and } \quad \psi = \frac{1}{2C}.$$

Recall some notations

$$A_0 = 0,\ A_1 = B_1,\ A_2 = A_1 + B_1,\ ..,\ A_N = A_{N-1} + B_N = \sum_{i=1}^{N} B_i.$$

Denote

$$u_0 = \frac{n}{\psi},\ u_1 = u_0 + \Delta,\ u_2 = u_1 + \Delta,\ ..,\ u_N = u_{N-1} + \Delta = 2CN + 2Cn.$$

Define

$$\psi_{u_i}(A_i) = \mathrm{Tr}\left(u_i.I_n - A_i\right)^{-1},$$

the corresponding potential function when $A_i \prec u_i.I_n$.
Denote by $\Im_i$ the event

$$\Im_i := \text{"}A_i \prec u_i.I_n \quad \text{ and } \quad \psi_{u_i}(A_i) \leqslant \psi\text{"}.$$

Clearly $\mathbb{P}\left(\Im_0\right) = 1$. Suppose now that $\Im_i$ is satisfied ; as we have seen in Lemma 4.19, the following condition is sufficient for the occurrence of the event $\Im_{i+1}$:

$$Q_2(\Delta, B_{i+1}) + Q_1(\Delta, B_{i+1}) \leqslant 1$$

Note that

$$Q_2(\Delta, B_{i+1}) \leqslant \frac{1}{\Delta} P_2(\Delta, B_{i+1}),$$

where P_2 is defined in (4.5).

Now denoting λ_j the eigenvalues of A_i and v_j the corresponding eigenvectors, taking the probability with respect to B_{i+1} one can write

$$\begin{aligned}
&\mathbb{P}\left(Q_2(\Delta, B_{i+1}) + Q_1(\Delta, B_{i+1}) > 1\right) \leqslant \mathbb{P}\left(\frac{1}{\Delta}P_2(\Delta, B_{i+1}) + Q_1(\Delta, B_{i+1}) > 1\right) \\
&\leqslant \mathbb{P}\left(\frac{1}{\Delta}P_2(\Delta, B_{i+1}) > \frac{1}{2}\right) + \mathbb{P}\left(Q_1(\Delta, B_{i+1}) > \frac{1}{2}\right) \\
&\leqslant \mathbb{P}\left(\sum_{j=1}^{n} \frac{\langle B_{i+1}v_j, v_j\rangle}{(u_{i+1}-\lambda_j)(u_i-\lambda_j)} > \frac{\Delta}{2}\sum_{j=1}^{n}\frac{1}{(u_{i+1}-\lambda_j)(u_i-\lambda_j)}\right) \\
&\quad + \mathbb{P}\left(\sum_{j=1}^{n}\frac{\langle B_{i+1}v_j, v_j\rangle}{u_{i+1}-\lambda_j} > \frac{1}{2}\right) \\
&\leqslant \mathbb{P}\left(\exists j \leqslant n/ \ \ \langle B_{i+1}v_j, v_j\rangle > C\right) + \mathbb{P}\left(\exists j \leqslant n/ \ \ \langle B_{i+1}v_j, v_j\rangle > \frac{1}{2\psi}\right) \\
&\leqslant 2n.\exp(-c\alpha)
\end{aligned}$$

So we have shown that $\mathbb{P}\left(\Im_{i+1}|\Im_i\right) \geqslant 1 - 2n.\exp(-c\alpha)$. Since B_i are independent we have

$$\begin{aligned}
\mathbb{P}\left(\lambda_{\max}\left(\frac{1}{N}\sum_{i=1}^{N}B_i\right) \leqslant 2C.(1+\frac{n}{N})\right) &\geqslant \mathbb{P}\left(\Im_N\right) \\
&\geqslant \mathbb{P}\left(\Im_N|\Im_{N-1}\right)\mathbb{P}\left(\Im_{N-1}|\Im_{N-2}\right)..\mathbb{P}\left(\Im_0\right) \\
&\geqslant 1 - 2Nn.\exp(-c\alpha)
\end{aligned}$$

Proposition 5.22 follows by (5.4). □

By the same techniques, we also get an estimate of the smallest eigenvalue.

Proposition 5.23. *Let n, m and N some fixed integers. Let $r > 2$ and A be an $n \times m$ isotropic $(-\frac{1}{r})$-concave matrix and denote $B = AA^t$. If*

$$\min\left\{r, \sqrt{m}\right\} \geqslant C\log(2nN), \tag{5.5}$$

then with probability $\geqslant 1 - \exp\left(-c\min\{r, \sqrt{m}\}\right)$ we have

$$\lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^{N}B_i\right) \geqslant c - C\frac{n}{N},$$

where $(B_i)_{i\leqslant N}$ are independent copies of B and c, C are universal positive constants.

Proof. Denote $\alpha = \min\left\{r, \sqrt{m}\right\}$. Here we will need the large deviation inequality and the small ball probability estimate satisfied by B. Applying Proposition 5.19 for rank 1 projections, we have

$$\mathbb{P}\left(\langle Bx, x\rangle \geqslant C\right) \leqslant \exp(-c\alpha) \quad \forall x \in \mathbb{S}^{n-1}.$$

Now applying Proposition 5.20 for rank 1 projections, we have

$$\mathbb{P}\left(\langle Bx, x\rangle \leqslant c\right) \leqslant \exp(-c\alpha) \quad \forall x \in \mathbb{S}^{n-1}.$$

Define δ and ϕ as follows :

$$\delta = \frac{c}{2} \quad \text{ and } \quad \phi = \frac{1}{C}$$

Recall some notations :
$A_0 = 0,\ A_1 = B_1,\ A_2 = A_1 + B_1,\ ..,\ A_N = A_{N-1} + B_N = \sum_{i=1}^{N} B_i$. Denote $l_0 = -\frac{n}{\phi},\ l_1 = l_0 + \delta,\ l_2 = l_1 + \delta,\ ..,\ l_N = l_{N-1} + \delta = \frac{c}{2}N - Cn$. Define $\phi_{l_i}(A_i) = \mathrm{Tr}\left(A_i - l_i.I_n\right)^{-1}$ the corresponding potential function when $A_i \succeq l_i.I_n$. Note also that $\delta \leqslant \frac{1}{\phi}$.
Denote by $\Im_i$ the event

$$\Im_i := \text{“}A_i \succeq l_i.I_n \quad \text{and} \quad \phi_{l_i}(A_i) \leqslant \phi\text{”}.$$

Clearly $\mathbb{P}\left(\Im_0\right) = 1$. Suppose now that $\Im_i$ is satisfied, following what was done after Lemma 4.15, condition (4.2) is sufficient for the occurrence of the event $\Im_{i+1}$:

$$\frac{1}{\delta} q_2(\delta, B_{i+1}) - q_1(\delta, B_{i+1}) \geqslant 1$$

Denoting λ_j the eigenvalues of A_i and v_j the corresponding eigenvectors, taking the probability with respect to B_{i+1} one can write

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{\delta} q_2(\delta, B_{i+1}) - q_1(\delta, B_{i+1}) < 1\right) \leqslant \\
&\leqslant \mathbb{P}\left(\frac{1}{\delta} q_2(\delta, B_{i+1}) < 2\right) + \mathbb{P}\left(q_1(\delta, B_{i+1}) > 1\right) \\
&\leqslant \mathbb{P}\left(\sum_{j=1}^{n} \frac{\langle B_{i+1} v_j, v_j\rangle}{(\lambda_j - l_{i+1})^2} < 2\delta \sum_{j=1}^{n} \frac{1}{(\lambda_j - l_{i+1})^2}\right) + \mathbb{P}\left(\sum_{j=1}^{n} \frac{\langle B_{i+1} v_j, v_j\rangle}{\lambda_j - l_{i+1}} > 1\right) \\
&\leqslant \mathbb{P}\left(\exists j \leqslant n /\ \langle B_{i+1} v_j, v_j\rangle < c\right) + \mathbb{P}\left(\exists j \leqslant n /\ \langle B_{i+1} v_j, v_j\rangle > \frac{1}{\phi}\right) \\
&\leqslant 2n.\exp(-c\alpha)
\end{aligned}$$

So we have shown that $\mathbb{P}\left(\Im_{i+1}|\Im_i\right) \geqslant 1 - 2n.\exp(-c\sqrt{m})$. Since B_i are independent we have :

$$\begin{aligned}
\mathbb{P}\left(\lambda_{\min}\left(\frac{1}{N}\sum_{i=1}^{N} B_i\right) \geqslant \frac{c}{2} - C\frac{n}{N}\right) &\geqslant \mathbb{P}\left(\Im_N\right) \\
&\geqslant \mathbb{P}\left(\Im_N|\Im_{N-1}\right)\mathbb{P}\left(\Im_{N-1}|\Im_{N-2}\right)..\mathbb{P}\left(\Im_0\right) \\
&\geqslant 1 - 2nN.\exp(-c\alpha)
\end{aligned}$$

Proposition 5.23 follows by (5.5).

□

Remark 5.24. *In a similar way to what is done in section 5.3, one can prove that taking A an $n \times m$ random matrix whose density with respect to Lebesgue is given by*

$$G(A) = \left(f(s_1(A), ..., s_k(A))\right)^{-(nm+r)},$$

where f is an absolutely symmetric convex function, properly normalized and $k = \min(n, m)$, then A is an isotropic $(-\frac{1}{r})$-concave matrix. Therefore, the results of this section apply to this class of matrices.

Bibliographie

[1] R. Adamczak, O. Guédon, R. Latała, K. Oleszkiewicz, A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Moment estimates for convex measures. *Electronic Journal of Probability*, 17(101) :1–19, 2012. 30, 61, 141, 142, 143, 155

[2] R. Adamczak, R. Latała, K. Oleszkiewicz, A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann. A short proof of Paouris' inequality. To appear in Can. Math. Bul. 30, 61, 142

[3] R. Adamczak, A. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Tail estimates for norms of sums of log-concave random vectors. arXiv :1107.4070v1, July 2011. 24, 55

[4] R. Adamczak, A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc.*, 23(2) :535–561, 2010. 24, 28, 55, 56, 59, 119

[5] R. Adamczak, A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Sharp bounds on the rate of convergence of the empirical covariance matrix. *C. R. Math. Acad. Sci. Paris*, 349(3-4) :195–200, 2011. 24, 55, 119

[6] R. Ahlswede and A. Winter. Addendum to : "Strong converse for identification via quantum channels" [IEEE Trans. Inform. Theory **48** (2002), no. 3, 569–579 ; MR1889969 (2003d :94069)]. *IEEE Trans. Inform. Theory*, 49(1) :346, 2003. 25, 56, 119

[7] J. Anderson. Extreme points in sets of positive linear maps on $\mathcal{B}(\mathcal{H})$. *J. Funct. Anal.*, 31(2) :195–217, 1979. 15, 47, 101

[8] G. Aubrun. Sampling convex bodies : a random matrix approach. *Proc. Amer. Math. Soc.*, 135(5) :1293–1303 (electronic), 2007. 120

[9] Z. D. Bai and Y. Q. Yin. Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.*, 21(3) :1275–1294, 1993. 27, 58

[10] K. Ball. Logarithmically concave functions and sections of convex sets in $\mathbf{R}^n$. *Studia Math.*, 88(1) :69–84, 1988. 28, 59

[11] K. Ball. Ellipsoids of maximal volume in convex bodies. *Geom. Dedicata*, 41(2) :241–250, 1992.

[12] J. Bastero and M. Romance. John's decomposition of the identity in the non-convex case. *Positivity*, 6(1) :1–16, 2002. 2, 34

[13] J. D. Batson, D. A. Spielman, and N. Srivastava. Twice-Ramanujan sparsifiers. In *STOC'09—Proceedings of the 2009 ACM International Symposium on Theory of Computing*, pages 255–262. ACM, New York, 2009. 1, 2, 6, 10, 11, 23, 25, 27, 33, 34, 38, 41, 43, 55, 56, 58, 66, 77, 81, 86, 87, 97, 100, 120, 130, 135

[14] K. Berman, H. Halpern, V. Kaftal, and G. Weiss. Matrix norm inequalities and the relative Dixmier property. *Integral Equations Operator Theory*, 11(1) :28–48, 1988. 11, 42, 80

[15] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997. 119

[16] C. Borell. Convex measures on locally convex spaces. *Ark. Mat.*, 12 :239–252, 1974. 28, 59, 141

[17] C. Borell. Convex set functions in d-space. *Period. Math. Hungar.*, 6(2) :111–136, 1975. 28, 59, 141

[18] J. Bourgain and S. J. Szarek. The Banach-Mazur distance to the cube and the Dvoretzky-Rogers factorization. *Israel J. Math.*, 62(2) :169–180, 1988. 19, 20, 51, 106, 107, 113, 114

[19] J. Bourgain and L. Tzafriri. Invertibility of "large" submatrices with applications to the geometry of Banach spaces and harmonic analysis. *Israel J. Math.*, 57(2) :137–224, 1987. 5, 10, 14, 16, 17, 36, 42, 46, 48, 64, 65, 80, 87, 97, 100, 101, 102, 103

[20] J. Bourgain and L. Tzafriri. On a problem of Kadison and Singer. *J. Reine Angew. Math.*, 420 :1–43, 1991. 10, 14, 15, 42, 46, 47

[21] P. Casazza. The simplified version of the Spielman and Srivastava algorithm for proving the Bourgain-Tzafriri restricted invertiblity theorem. Available at arXiv :1208.4013. 73, 74

[22] P. G. Casazza, O. Christensen, A. M. Lindner, and R. Vershynin. Frames and the Feichtinger conjecture. *Proc. Amer. Math. Soc.*, 133(4) :1025–1033 (electronic), 2005. 15, 47

[23] P. G. Casazza and D. Edidin. Equivalents of the Kadison-Singer problem. In *Function spaces*, volume 435 of *Contemp. Math.*, pages 123–142. Amer. Math. Soc., Providence, RI, 2007. 15, 47, 87, 98, 100, 101

[24] P. G. Casazza and J. C. Tremain. Revisiting the Bourgain-Tzafriri restricted invertibility theorem. *Oper. Matrices*, 3(1) :97–110, 2009.

[25] D. Chafai, O. Guédon, G. Lecué, and A. Pajor. Interactions between compressed sensing, random matrices and high dimensional geometry. 122

[26] M. De Carli Silva, N. Harvey, and C. Sato. Sparse Sums of Positive Semidefinite Matrices. arXiv :1107.0088v2, July 2011. 27, 58, 120, 130, 135

[27] J. Diestel, H. Jarchow, and A. Tonge. *Absolutely summing operators*, volume 43 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995. 5, 9, 37, 41, 65, 77, 81, 98, 107

[28] A. Dvoretzky and C. A. Rogers. Absolute and unconditional convergence in normed linear spaces. *Proc. Nat. Acad. Sci. U. S. A.*, 36 :192–197, 1950. 20, 51, 106

[29] M. Fradelizi. Concentration inequalities for s-concave measures of dilations of Borel sets and applications. *Electron. J. Probab.*, 14 :no. 71, 2068–2090, 2009.

[30] A. Giannopoulos, I. Perissinaki, and A. Tsolomitis. John's theorem for an arbitrary pair of convex bodies. *Geom. Dedicata*, 84(1-3) :63–79, 2001. 2, 34

[31] A. A. Giannopoulos. A note on the Banach-Mazur distance to the cube. In *Geometric aspects of functional analysis (Israel, 1992–1994)*, volume 77 of *Oper. Theory Adv. Appl.*, pages 67–73. Birkhäuser, Basel, 1995. 19, 21, 51, 52, 107, 113, 114, 116

[32] A. A. Giannopoulos. A proportional Dvoretzky-Rogers factorization result. *Proc. Amer. Math. Soc.*, 124(1) :233–241, 1996. 21, 52, 107

[33] E. Gluskin and A. Litvak. A remark on vertex index of the convex bodies. In B. Klartag, S. Mendelson, and V. D. Milman, editors, *Geometric Aspects of Functional Analysis*, volume 2050 of *Lecture Notes in Mathematics*, pages 255–265. Springer Berlin Heidelberg, 2012. 3, 35

[34] E. D. Gluskin. The diameter of the Minkowski compactum is roughly equal to n. *Funktsional. Anal. i Prilozhen.*, 15(1) :72–73, 1981. 19, 50

[35] Y. Gordon, A. E. Litvak, M. Meyer, and A. Pajor. John's decomposition in the general case and applications. *J. Differential Geom.*, 68(1) :99–119, 2004. 2, 34

[36] O. Guédon and E. Milman. Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures. *Geom. Funct. Anal.*, 21(5) :1043–1068, 2011. 29, 60, 143

[37] V. I. Gurariĭ, M. Ĭ. Kadec′, and V. I. Macaev. Distances between finite-dimensional analogs of the L_p-spaces. *Mat. Sb. (N.S.)*, 70 (112) :481–489, 1966. 18, 50

[38] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58 :13–30, 1963.

[39] R. A. Horn and C. R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1994. Corrected reprint of the 1991 original.

[40] F. John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, pages 187–204. Interscience Publishers, Inc., New York, N. Y., 1948. 2, 19, 34, 50, 95, 96, 108, 111, 116

[41] R. V. Kadison and I. M. Singer. Extensions of pure states. *Amer. J. Math.*, 81 :383–400, 1959. 15, 47, 87, 101

[42] R. Kannan, L. Lovász, and M. Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures Algorithms*, 11(1) :1–50, 1997. 119

[43] B. Kashin and L. Tzafriri. Some remarks on the restriction of operators to coordinate subspaces. preprint. 6, 9, 38, 40, 41, 65, 76, 97

[44] B. S. Kašin. Some properties of matrices of bounded operators from the space l_2^n into l_2^m. *Izv. Akad. Nauk Armyan. SSR Ser. Mat.*, 15(5) :379–394, 417, 1980. 76, 97

[45] V. F. Kolchin, B. A. Sevast′yanov, and V. P. Chistyakov. *Random allocations.* V. H. Winston & Sons, Washington, D.C., 1978. Translated from the Russian, Translation edited by A. V. Balakrishnan, Scripta Series in Mathematics. 120

[46] M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)].* Springer-Verlag, Berlin, 1991. Isoperimetry and processes.

[47] A. S. Lewis. The convex analysis of unitarily invariant matrix functions. *J. Convex Anal.*, 2(1-2) :173–183, 1995. 151

[48] D. R. Lewis. Finite dimensional subspaces of L_p. *Studia Math.*, 63(2) :207–212, 1978.

[49] D. R. Lewis. Ellipsoids defined by Banach ideal norms. *Mathematika*, 26(1) :18–29, 1979. 2, 34

[50] A. E. Litvak and N. Tomczak-Jaegermann. Random aspects of high-dimensional convex bodies. In *Geometric aspects of functional analysis*, volume 1745 of *Lecture Notes in Math.*, pages 169–190. Springer, Berlin, 2000. 22, 53, 107, 111

[51] A. A. Lunin. On operator norms of submatrices. *Mat. Zametki*, 45(3) :94–100, 128, 1989. 9, 40, 76, 97

[52] F. Lust-Piquard. Inégalités de Khintchine dans C_p $(1 < p < \infty)$. *C. R. Acad. Sci. Paris Sér. I Math.*, 303(7) :289–292, 1986. 26, 57

[53] F. Lust-Piquard and G. Pisier. Noncommutative Khintchine and Paley inequalities. *Ark. Mat.*, 29(2) :241–260, 1991. 26, 57

[54] L. Mackey, M. Jordan, R. Chen, B. Farrell, and J. Tropp. Matrix concentration inequalities via the method of exchangeable pairs. Available at arXiv :1201.6002. 25, 56, 119

[55] A. Naor. Sparse quadratic forms and their geometric applications (after Batson, Spielman and Srivastava). Available at arXiv :1101.4324. 11, 43, 81, 87, 100

[56] D. Needell and J. Tropp. Paved with good intentions : Analysis of a randomized block kaczmarz method. Available at arXiv :1208.3805.

[57] R. I. Oliveira. Sums of random Hermitian matrices and an inequality by Rudelson. *Electron. Commun. Probab.*, 15 :203–212, 2010. 26, 57, 120

[58] G. Paouris. Concentration of mass on convex bodies. *Geom. Funct. Anal.*, 16(5) :1021–1049, 2006. 28, 59

[59] G. Paouris. Concentration of mass on isotropic convex bodies. *C. R. Math. Acad. Sci. Paris*, 342(3) :179–182, 2006. 28, 59, 142

[60] G. Paouris. Small ball probability estimates for log-concave measures. *Trans. Amer. Math. Soc.*, 364(1) :287–308, 2012. 28, 59, 142

[61] G. Pisier. *Factorization of linear operators and geometry of Banach spaces*, volume 60 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1986. 5, 9, 37, 41, 65, 77, 81, 98, 107

[62] G. Pisier. *The volume of convex bodies and Banach space geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989.

[63] H. P. Rosenthal. On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables. *Israel J. Math.*, 8 :273–303, 1970. 123

[64] M. Rudelson. Approximate John's decompositions. In *Geometric aspects of functional analysis (Israel, 1992–1994)*, volume 77 of *Oper. Theory Adv. Appl.*, pages 245–249. Birkhäuser, Basel, 1995.

[65] M. Rudelson. Contact points of convex bodies. *Israel J. Math.*, 101 :93–124, 1997. 3, 34

[66] M. Rudelson. Almost orthogonal submatrices of an orthogonal matrix. *Israel J. Math.*, 111 :143–155, 1999.

[67] M. Rudelson. Random vectors in the isotropic position. *J. Funct. Anal.*, 164(1) :60–72, 1999. 24, 55, 119

[68] M. Rudelson and R. Vershynin. Sampling from large matrices : an approach through geometric functional analysis. *J. ACM*, 54(4) :Art. 21, 19 pp. (electronic), 2007.

[69] I. Z. Ruzsa. On difference sets. *Studia Sci. Math. Hungar.*, 13(3-4) :319–326 (1981), 1978. 103

[70] N. Sauer. On the density of families of sets. *J. Combinatorial Theory Ser. A*, 13 :145–147, 1972. 5, 37, 65

[71] G. Schechtman. Dimension reduction in l_p, $0 < p < 2$. Available at arXiv :1110.2148. 3, 35

[72] G. Schechtman. Tight embedding of subspaces of L_p in ℓ_p^n for even p. *Proc. Amer. Math. Soc.*, 139(12) :4419–4421, 2011. 3, 35

[73] S. Shelah. A combinatorial problem ; stability and order for models and theories in infinitary languages. *Pacific J. Math.*, 41 :247–261, 1972. 5, 37, 65

[74] D. A. Spielman and N. Srivastava. An elementary proof of the restricted invertibility theorem. *Israel J. Math.*, 190 :83–91, 2012. 6, 38, 66

[75] D. A. Spielman and S.-H. Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 81–90 (electronic), New York, 2004. ACM. 2, 34

[76] D. A. Spielman and S.-H. Teng. Spectral sparsification of graphs. *SIAM J. Comput.*, 40(4) :981–1025, 2011. 2, 34

[77] N. Srivastava. On Contact Points of Convex Bodies. to appear in Geometric Aspects of Functional Analysis, Springer Lec. Notes in Math. 2, 34

[78] N. Srivastava. *Spectral Sparsification and Restricted Invertibility*. PhD thesis, Yale University, Mar. 2010. 69, 78

[79] N. Srivastava and R. Vershynin. Covariance Estimation for Distributions with $(2+\epsilon)$-Moments. to appear in Annals of Probability. 25, 56, 119, 120, 121, 136, 145

[80] S. J. Szarek. Spaces with large distance to l_∞^n and random matrices. *Amer. J. Math.*, 112(6) :899–942, 1990. 19, 20, 50, 51, 52, 107, 113

[81] S. J. Szarek. On the geometry of the Banach-Mazur compactum. In *Functional analysis (Austin, TX, 1987/1989)*, volume 1470 of *Lecture Notes in Math.*, pages 48–59. Springer, Berlin, 1991.

[82] S. J. Szarek and M. Talagrand. An "isomorphic" version of the Sauer-Shelah lemma and the Banach-Mazur distance to the cube. In *Geometric aspects of functional analysis (1987–88)*, volume 1376 of *Lecture Notes in Math.*, pages 105–112. Springer, Berlin, 1989. 19, 20, 51, 52, 107, 113, 114

[83] M. Talagrand. Embedding of l_k^∞ and a theorem of Alon and Milman. In *Geometric aspects of functional analysis (Israel, 1992–1994)*, volume 77 of *Oper. Theory Adv. Appl.*, pages 289–293. Birkhäuser, Basel, 1995. 12, 44

[84] S. Taschuk. The Banach-Mazur distance to the cube in low dimensions. *Discrete Comput. Geom.*, 46(1) :175–183, 2011. 19, 51, 117

[85] N. Tomczak-Jaegermann. *Banach-Mazur distances and finite-dimensional operator ideals*, volume 38 of *Pitman Monographs and Surveys in Pure and Applied Mathematics*. Longman Scientific & Technical, Harlow, 1989. 97

[86] J. A. Tropp. The random paving property for uniformly bounded matrices. *Studia Math.*, 185(1) :67–82, 2008. 14, 46, 101

[87] J. A. Tropp. Column subset selection, matrix factorization, and eigenvalue optimization. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 978–986, Philadelphia, PA, 2009. SIAM. 5, 9, 37, 41, 65, 77, 87, 97, 100

[88] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4) :389–434, 2012. 25, 56, 119, 125

[89] R. Vershynin. Coordinate restrictions of linear operators in l_2^n. preprint. 17, 48, 102, 103

[90] R. Vershynin. John's decompositions : selecting a large part. *Israel J. Math.*, 122 :253–277, 2001. 5, 9, 12, 37, 40, 41, 43, 65, 77, 85, 95

[91] R. Vershynin. Approximating the moments of marginals of high-dimensional distributions. *Ann. Probab.*, 39(4) :1591–1606, 2011.

[92] R. Vershynin. How Close is the Sample Covariance Matrix to the Actual Covariance Matrix ? *J. Theoret. Probab.*, 25(3) :655–686, 2012.

[93] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In *Compressed sensing*, pages 210–268. Cambridge Univ. Press, Cambridge, 2012.

[94] P. Youssef. Estimating the covariance of random matrices. Available at arXiv :1301.6607.

[95] P. Youssef. A note on column subset selection. Available at arXiv :1212.0976.

[96] P. Youssef. Restricted invertibility and the Banach-Mazur distance to the cube. Available at arXiv :1206.0654.